

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
10 January 2002 (10.01.2002)

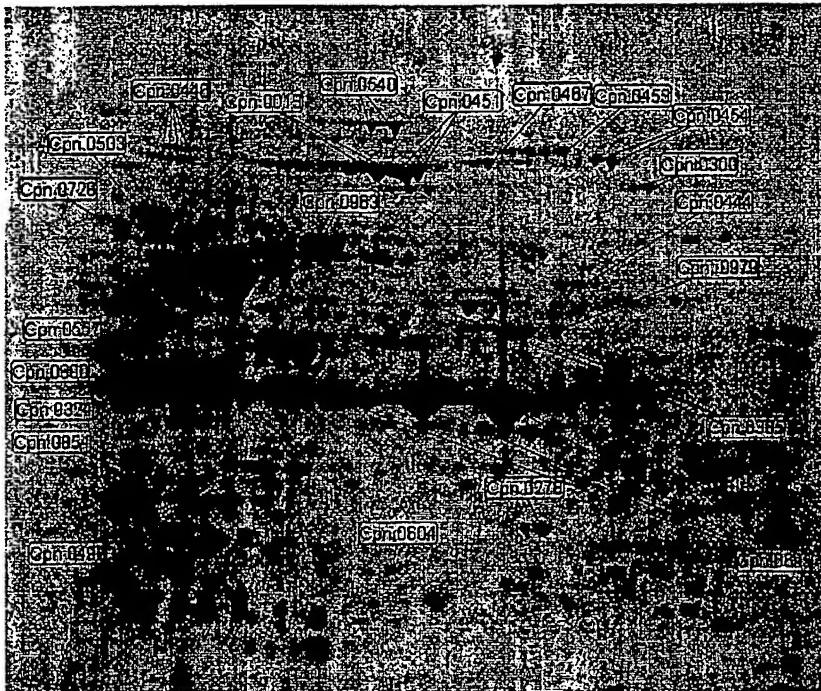
PCT

(10) International Publication Number
WO 02/02606 A2

- (51) International Patent Classification⁷: C07K 14/295, C12N 15/31, A61K 39/118
- (21) International Application Number: PCT/IB01/01445
- (22) International Filing Date: 3 July 2001 (03.07.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
- | | | |
|-----------|--------------------------------|----|
| 0016363.4 | 3 July 2000 (03.07.2000) | GB |
| 0017047.2 | 11 July 2000 (11.07.2000) | GB |
| 0017983.8 | 21 July 2000 (21.07.2000) | GB |
| 0019368.0 | 7 August 2000 (07.08.2000) | GB |
| 0020440.4 | 18 August 2000 (18.08.2000) | GB |
| 0022583.9 | 14 September 2000 (14.09.2000) | GB |
| 0027549.5 | 10 November 2000 (10.11.2000) | GB |
| 0031706.5 | 22 December 2000 (22.12.2000) | GB |
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): RATTI, Giulio [IT/IT]; Chiron S.p.A., Via Fiorentina, 1, I-53100 Siena (IT). GRANDI, Guido [IT/IT]; Chiron S.p.A., Via Fiorentina, 1, I-53100 Siena (IT).
- (74) Agents: HALLYBONE, Huw, George et al.; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European

[Continued on next page]

(54) Title: IMMUNISATION AGAINST *CHLAMYDIA PNEUMONIAE*



(57) Abstract: The published genomic of *Chlamydia pneumoniae* reveals over 1000 putative encoded proteins but does not itself indicate which of these might be useful antigens for immunisation and vaccination or for diagnosis. This difficulty is addressed by the invention, which provides a number of *C. pneumoniae* protein sequences suitable for vaccine production and development and/or for diagnostic purposes.



patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Published:

- *without international search report and to be republished upon receipt of that report*

IMMUNISATION AGAINST *CHLAMYDIA PNEUMONIAE*

All documents cited herein are incorporated by reference in their entirety.

TECHNICAL FIELD

This invention is in the field of immunisation against chlamydial infection, in particular against
5 infection by *Chlamydia pneumoniae*.

BACKGROUND ART

Chlamydiae are obligate intracellular parasites of eukaryotic cells which are responsible for endemic sexually transmitted infections and various other disease syndromes. They occupy an exclusive eubacterial phylogenetic branch, having no close relationship to any other known organisms – they are
10 classified in their own order (*Chlamydiales*) which contains a single family (*Chlamydiaceae*) which in turn contains a single genus (*Chlamydia*). A particular characteristic of the *Chlamydiae* is their unique life cycle, in which the bacterium alternates between two morphologically distinct forms: an extracellular infective form (elementary bodies, EB) and an intracellular non-infective form (reticulate bodies, RB). The life cycle is completed with the re-organization of RB into EB, which
15 subsequently leave the disrupted host cell ready to infect further cells.

Four chlamydial species are currently known – *C.trachomatis*, *C.pneumoniae*, *C.pecorum* and
15 *C.psittaci* [e.g. Raulston (1995) *Mol Microbiol* 15:607-616; Everett (2000) *Vet Microbiol* 75:109-126]. *C.pneumoniae* is closely related to *C.trachomatis*, as the whole genome comparison of at least two isolates from each species has shown [Kalman *et al.* (1999) *Nature Genetics* 21:385-389; Read
20 *et al.* (2000) *Nucleic Acids Res* 28:1397-406; Stephens *et al.* (1998) *Science* 282:754-759]. Based on surface reaction with patient immune sera, the current view is that only one serotype of *C.pneumoniae* exists world-wide.

C.pneumoniae is a common cause of human respiratory disease. It was first isolated from the conjunctiva of a child in Taiwan in 1965, and was established as a major respiratory pathogen in
25 1983. In the USA, *C.pneumoniae* causes approximately 10% of community-acquired pneumonia and 5% of pharyngitis, bronchitis, and sinusitis.

More recently, the spectrum of *C.pneumoniae* infections has been extended to include atherosclerosis, coronary heart disease, carotid artery stenosis, myocardial infarction, cerebrovascular disease, aortic aneurysm, claudication, and stroke. The association of *C.pneumoniae* with
30 atherosclerosis is corroborated by the presence of the organism in atherosclerotic lesions throughout the arterial tree and the near absence of the organism in healthy arterial tissue. *C.pneumoniae* has also been isolated from coronary and carotid atheromatous plaques. The bacterium has also been associated with other acute and chronic respiratory diseases (e.g. otitis media, chronic obstructive pulmonary disease, pulmonary exacerbation of cystic fibrosis) as a result of sero-epidemiologic
35 observations, case reports, isolation or direct detection of the organism in specimens, and successful

response to anti-chlamydial antibiotics. To determine whether chronic infection plays a role in initiation or progression of disease, intervention studies in humans have been initiated, and animal models of *C.pneumoniae* infection have been developed.

- Considerable knowledge of the epidemiology of *C.pneumoniae* infection has been derived from
5 serologic studies using the *C.pneumoniae*-specific microimmunofluorescence test. Infection is ubiquitous, and it is estimated that virtually everyone is infected at some point in life, with common re-infection. Antibodies against *C.pneumoniae* are rare in children under the age of 5, except in developing and tropical countries. Antibody prevalence increases rapidly at ages 5 to 14, reaching 50% at the age of 20, and continuing to increase slowly to ~80% by age 70.
- 10 A current hypothesis is that *C.pneumoniae* can persist in an asymptomatic low-grade infection in very large sections of the human population. When this condition occurs, it is believed that the presence of *C.pneumoniae*, and/or the effects of the host reaction to the bacterium, can cause or help progress of cardiovascular illness.

- It is not yet clear whether *C.pneumoniae* is actually a causative agent of cardiovascular disease, or
15 whether it is just artefactually associated with it. It has been shown, however, that *C.pneumoniae* infection can induce LDL oxidation by human monocytes [Kalayoglu *et al.* (1999) *J. Infect. Dis.* 180:780-90; Kalayoglu *et al.* (1999) *Am. Heart J.* 138:S488-490]. As LDL oxidation products are highly atherogenic, this observation provides a possible mechanism whereby *C.pneumoniae* may cause atherosomatous degeneration. If a causative effect is confirmed, vaccination (prophylactic and
20 therapeutic) will be universally recommended.

- Genomic sequence information has been published for *C.pneumoniae* [Kalman *et al.* (1999) *supra*; Read *et al.* (2000) *supra*; Shirai *et al.* (2000) *J. Infect. Dis.* 181(Suppl 3):S524-S527; WO99/27105; WO00/27994] and is available from GenBank. Sequencing efforts have not, however, focused on vaccination, and the availability of genomic sequence does not in itself indicate which of the >1000 genes might encode useful antigens for immunisation and vaccination. WO99/27105, for instance, implies that every one of the 1296 ORFs identified in the *C.pneumoniae* strain CM1 genome is a useful vaccine antigen.

- It is thus an object of the present invention to identify antigens useful for vaccine production and development from amongst the many proteins present in *C.pneumoniae*. It is a further object to
30 identify antigens useful for diagnosis (*e.g.* immunodiagnosis) of *C.pneumoniae*.

DISCLOSURE OF THE INVENTION

The invention provides proteins comprising the *C.pneumoniae* amino acid sequences disclosed in the examples.

- It also provides proteins comprising sequences which share at least $x\%$ sequence identity with the
35 *C.pneumoniae* amino acid sequences disclosed in the examples. Depending on the particular

sequence, x is preferably 50% or more (e.g. 60%, 70%, 80%, 90%, 95%, 99% or more). These include mutants and allelic variants. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between proteins is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH program (Oxford Molecular), using an affine gap search with parameters *gap open penalty*=12 and *gap extension penalty*=1.

The invention further provides proteins comprising fragments of the *C.pneumoniae* amino acid sequences disclosed in the examples. The fragments should comprise at least n consecutive amino acids from the sequences and, depending on the particular sequence, n is 7 or more (e.g. 8, 10, 12, 10 14, 16, 18, 20, 30, 40, 50, 75, 100 or more). Preferably the fragments comprise one or more epitope(s) from the sequence. Other preferred fragments omit a signal peptide.

The proteins of the invention can, of course, be prepared by various means (e.g. native expression, recombinant expression, purification from cell culture, chemical synthesis etc.) and in various forms (e.g. native, fusions etc.). They are preferably prepared in substantially pure form (ie. substantially free from other *C.pneumoniae* or host cell proteins). Heterologous expression in *E.coli* is a preferred preparative route.

According to a further aspect, the invention provides nucleic acid comprising the *C.pneumoniae* nucleotide sequences disclosed in the examples. In addition, the invention provides nucleic acid comprising sequences which share at least $x\%$ sequence identity with the *C.pneumoniae* nucleotide sequences disclosed in the examples. Depending on the particular sequence, x is preferably 50% or more (e.g. 60%, 70%, 80%, 90%, 95%, 99% or more).

Furthermore, the invention provides nucleic acid which can hybridise to the *C.pneumoniae* nucleic acid disclosed in the examples, preferably under "high stringency" conditions (e.g. 65°C in a 0.1xSSC, 0.5% SDS solution).

25 Nucleic acid comprising fragments of these sequences are also provided. These should comprise at least n consecutive nucleotides from the *C.pneumoniae* sequences and, depending on the particular sequence, n is 10 or more (e.g. 12, 14, 15, 18, 20, 25, 30, 35, 40, 50, 75, 100, 200, 300 or more).

According to a further aspect, the invention provides nucleic acid encoding the proteins and protein fragments of the invention.

30 It should also be appreciated that the invention provides nucleic acid comprising sequences complementary to those described above (e.g. for antisense or probing purposes).

Nucleic acid according to the invention can, of course, be prepared in many ways (e.g. by chemical synthesis, from genomic or cDNA libraries, from the organism itself etc.) and can take various forms (e.g. single stranded, double stranded, vectors, probes etc.).

In addition, the term "nucleic acid" includes DNA and RNA, and also their analogues, such as those containing modified backbones, and also peptide nucleic acids (PNA) etc.

According to a further aspect, the invention provides vectors comprising nucleotide sequences of the invention (e.g. cloning or expression vectors) and host cells transformed therewith.

- 5 According to a further aspect, the invention provides immunogenic compositions comprising protein and/or nucleic acid according to the invention. These compositions are suitable for immunisation and vaccination purposes. Vaccines of the invention may be prophylactic or therapeutic, and will typically comprise an antigen which can induce antibodies capable of inhibiting (a) chlamydial adhesion, (b) chlamydial entry, and/or (c) successful replication within the host cell. The vaccines
10 preferably induce any cell-mediated T-cell responses which are necessary for chlamydial clearance from the host.

The invention also provides nucleic acid or protein according to the invention for use as medicaments (e.g. as vaccines). It also provides the use of nucleic acid or protein according to the invention in the manufacture of a medicament (e.g. a vaccine or an immunogenic composition) for
15 treating or preventing infection due to *C.pneumoniae*.

The invention also provides a method of treating (e.g. immunising) a patient, comprising administering to the patient a therapeutically effective amount of nucleic acid or protein according to the invention.

According to further aspects, the invention provides various processes.

- 20 A process for producing proteins of the invention is provided, comprising the step of culturing a host cell according to the invention under conditions which induce protein expression.

A process for producing protein or nucleic acid of the invention is provided, wherein the protein or nucleic acid is synthesised in part or in whole using chemical means.

- 25 A process for detecting *C.pneumoniae* in a sample is provided, wherein the sample is contacted with an antibody which binds to a protein of the invention .

A summary of standard techniques and procedures which may be employed in order to perform the invention (e.g. to utilise the disclosed sequences for immunisation) follows. This summary is not a limitation on the invention but, rather, gives examples that may be used, but are not required.

General

- 30 The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature e.g. Sambrook *Molecular Cloning; A Laboratory Manual*, Second Edition (1989) and Third Edition (2001); *DNA Cloning, Volumes I and ii* (D.N. Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); *Transcription and Translation* (B.D. Hames & S.J. Higgins eds. 1984); *Animal Cell Culture* (R.I.
35

Freshney ed. 1986); *Immobilized Cells and Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide to Molecular Cloning* (1984); the *Methods in Enzymology* series (Academic Press, Inc.), especially volumes 154 & 155; *Gene Transfer Vectors for Mammalian Cells* (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Mayer and Walker, eds. (1987), *Immunochemical Methods in Cell and Molecular Biology* (Academic Press, London); Scopes, (1987) *Protein Purification: Principles and Practice*, Second Edition (Springer-Verlag, N.Y.), and *Handbook of Experimental Immunology, Volumes I-IV* (D.M. Weir and C. C. Blackwell eds 1986).

5 Standard abbreviations for nucleotides and amino acids are used in this specification.

Definitions

10 A composition containing X is "substantially free of" Y when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95% or even 99% by weight.

The term "comprising" means "including" as well as "consisting" e.g. a composition "comprising" X may consist exclusively of X or may include something additional to X, such as X+Y.

15 The term "heterologous" refers to two biological components that are not found together in nature. The components may be host cells, genes, or regulatory regions, such as promoters. Although the heterologous components are not found together in nature, they can function together, as when a promoter heterologous to a gene is operably linked to the gene. Another example is where a Chlamydial sequence is heterologous to a mouse host cell. A further examples would be two epitopes from the same or different proteins which have been
20 assembled in a single protein in an arrangement not found in nature.

An "origin of replication" is a polynucleotide sequence that initiates and regulates replication of polynucleotides, such as an expression vector. The origin of replication behaves as an autonomous unit of polynucleotide replication within a cell, capable of replication under its own control. An origin of replication may be needed for a vector to replicate in a particular host cell. With certain origins of replication, an expression vector can be reproduced at a high copy number in the presence of the appropriate proteins within the cell. Examples of origins are the autonomously replicating sequences, which are effective in yeast; and the viral T-antigen, effective in COS-7 cells.

25 A "mutant" sequence is defined as DNA, RNA or amino acid sequence differing from but having sequence identity with the native or disclosed sequence. Depending on the particular sequence, the degree of sequence identity between the native or disclosed sequence and the mutant sequence is preferably greater than 50% (e.g. 60%, 70%, 80%, 90%, 95%, 99% or more, calculated using the Smith-Waterman algorithm as described above). As used herein, an "allelic variant" of a nucleic acid molecule, or region, for which nucleic acid sequence is provided herein is a nucleic acid molecule, or region, that occurs essentially at the same locus in the genome of another or second isolate, and that, due to natural variation caused by, for example, mutation or recombination,
30 has a similar but not identical nucleic acid sequence. A coding region allelic variant typically encodes a protein having similar activity to that of the protein encoded by the gene to which it is being compared. An allelic variant can also comprise an alteration in the 5' or 3' untranslated regions of the gene, such as in regulatory control regions (e.g. see US patent 5,753,235).

Expression systems

The Chlamydial nucleotide sequences can be expressed in a variety of different expression systems; for example those used with mammalian cells, baculoviruses, plants, bacteria, and yeast.

i. Mammalian Systems

5 Mammalian expression systems are known in the art. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence (e.g. structural gene) into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, usually located 25-30 base pairs (bp) upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element, usually located within 100 to 200 bp upstream of the TATA box. An upstream promoter element determines the rate at 10 which transcription is initiated and can act in either orientation [Sambrook et al. (1989) "Expression of Cloned Genes in Mammalian Cells." In *Molecular Cloning: A Laboratory Manual*, 2nd ed.].

15 Mammalian viral genes are often highly expressed and have a broad host range; therefore sequences encoding mammalian viral genes provide particularly useful promoter sequences. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter (Ad MLP), and herpes simplex virus promoter. In addition, sequences derived from non-viral genes, such as the murine metallothionein gene, also provide useful promoter sequences. Expression may be either constitutive or regulated (inducible), depending on the promoter can be induced with glucocorticoid in hormone-responsive 20 cells.

25 The presence of an enhancer element (enhancer), combined with the promoter elements described above, will usually increase expression levels. An enhancer is a regulatory DNA sequence that can stimulate transcription up to 1000-fold when linked to homologous or heterologous promoters, with synthesis beginning at the normal RNA start site. Enhancers are also active when they are placed upstream or downstream from the transcription initiation site, in either normal or flipped orientation, or at a distance of more than 1000 nucleotides from the promoter [Maniatis et al. (1987) *Science* 236:1237; Alberts et al. (1989) *Molecular Biology of the Cell*, 2nd ed.]. Enhancer elements derived from viruses may be particularly useful, because they usually have a broader host range. Examples include the SV40 early gene enhancer [Dijkema et al. (1985) *EMBO J.* 4:761] and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus [Gorman et al. 30 (1982) *PNAS USA* 79:6777] and from human cytomegalovirus [Boshart et al. (1985) *Cell* 41:521]. Additionally, some enhancers are regulatable and become active only in the presence of an inducer, such as a hormone or metal ion [Sassone-Corsi and Borelli (1986) *Trends Genet.* 2:215; Maniatis et al. (1987) *Science* 236:1237].

35 A DNA molecule may be expressed intracellularly in mammalian cells. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always be a methionine, which is encoded by the ATG start codon. If desired, the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in mammalian cells. Preferably, there are processing sites encoded between the leader

fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The adenovirus tripartite leader is an example of a leader sequence that provides for secretion of a foreign protein in mammalian cells.

- 5 Usually, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-transcriptional cleavage and polyadenylation [Birnstiel et al. (1985) *Cell* 41:349; Proudfoot and Whitelaw (1988) "Termination and 3' end processing of eukaryotic RNA. In *Transcription and splicing* (ed. B.D. Hames and D.M. Glover); Proudfoot 10 (1989) *Trends Biochem. Sci.* 14:105]. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator/polyadenylation signals include those derived from SV40 [Sambrook et al (1989) "Expression of cloned genes in cultured mammalian cells." In *Molecular Cloning: A Laboratory Manual*].

- Usually, the above described components, comprising a promoter, polyadenylation signal, and transcription 15 termination sequence are put together into expression constructs. Enhancers, introns with functional splice donor and acceptor sites, and leader sequences may also be included in an expression construct, if desired. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (e.g. plasmids) capable of stable maintenance in a host, such as mammalian cells or bacteria. Mammalian replication systems include those derived from animal viruses, which require trans-acting factors to replicate. For example, plasmids containing 20 the replication systems of papovaviruses, such as SV40 [Gluzman (1981) *Cell* 23:175] or polyomavirus, replicate to extremely high copy number in the presence of the appropriate viral T antigen. Additional examples of mammalian replicons include those derived from bovine papillomavirus and Epstein-Barr virus. Additionally, the replicon may have two replicaton systems, thus allowing it to be maintained, for example, in mammalian 25 cells for expression and in a prokaryotic host for cloning and amplification. Examples of such mammalian-bacteria shuttle vectors include pMT2 [Kaufman et al. (1989) *Mol. Cell. Biol.* 9:946] and pHEBO [Shimizu et al. (1986) *Mol. Cell. Biol.* 6:1074].

- The transformation procedure used depends upon the host to be transformed. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene-mediated transfection, protoplast fusion, 30 electroporation, encapsulation of polynucleotide(s) in liposomes, direct microinjection of the DNA into nuclei.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (e.g. Hep G2), and a number of other cell lines.

35 ii. Baculovirus Systems

- The polynucleotide encoding the protein can also be inserted into a suitable insect expression vector, and is operably linked to the control elements within that vector. Vector construction employs techniques which are known in the art. Generally, the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site 40 for insertion of the heterologous gene or genes to be expressed; a wild type baculovirus with a sequence

homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media.

After inserting the DNA sequence encoding the protein into the transfer vector, the vector and the wild type viral

5 genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit). These techniques are generally known to those skilled in the art and fully described in Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) 10 (hereinafter "Summers and Smith").

Prior to inserting the DNA sequence encoding the protein into the baculovirus genome, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are usually assembled into an intermediate transplacement construct (transfer vector). This construct may contain a single gene and operably linked regulatory elements; multiple genes, each with its 15 owned set of operably linked regulatory elements; or multiple genes, regulated by the same set of regulatory elements. Intermediate transplacement constructs are often maintained in a replicon, such as an extrachromosomal element (e.g. plasmids) capable of stable maintenance in a host, such as a bacterium. The replicon will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification.

20 Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 basepairs downstream from the ATT; see Luckow and Summers, *Virology* (1989) 17:31.

The plasmid usually also contains the polyhedrin polyadenylation signal (Miller et al. (1988) *Ann. Rev. 25 Microbiol.*, 42:177) and a prokaryotic ampicillin-resistance (*amp*) gene and origin of replication for selection and propagation in *E. coli*.

Baculovirus transfer vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (5' to 3') transcription of a coding sequence (e.g. structural gene) into mRNA. A promoter will have a transcription initiation region 30 which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A baculovirus transfer vector may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Expression may be either regulated or constitutive.

Structural genes, abundantly transcribed at late times in a viral infection cycle, provide particularly useful 35 promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedron protein, Friesen et al., (1986) "The Regulation of Baculovirus Gene Expression," in: *The Molecular Biology of Baculoviruses* (ed. Walter Doerfler); EPO Publ. Nos. 127 839 and 155 476; and the gene encoding the p10 protein, Vlak et al., (1988), *J. Gen. Virol.* 69:765.

DNA encoding suitable signal sequences can be derived from genes for secreted insect or baculovirus proteins, 40 such as the baculovirus polyhedrin gene (Carbonell et al. (1988) *Gene*, 73:409). Alternatively, since the signals

for mammalian cell posttranslational modifications (such as signal peptide cleavage, proteolytic cleavage, and phosphorylation) appear to be recognized by insect cells, and the signals required for secretion and nuclear accumulation also appear to be conserved between the invertebrate cells and vertebrate cells, leaders of non-insect origin, such as those derived from genes encoding human α -interferon, Maeda et al., (1985), *Nature* 315:592; human gastrin-releasing peptide, Lebacq-Verheyden et al., (1988), *Molec. Cell. Biol.* 8:3129; human IL-2, Smith et al., (1985) *Proc. Nat'l Acad. Sci. USA*, 82:8404; mouse IL-3, (Miyajima et al., (1987) *Gene* 58:273; and human glucocerebrosidase, Martin et al., (1988) *DNA*, 7:99, can also be used to provide for secretion in insects.

A recombinant polypeptide or polyprotein may be expressed intracellularly or, if it is expressed with the proper regulatory sequences, it can be secreted. Good intracellular expression of nonfused foreign proteins usually requires heterologous genes that ideally have a short leader sequence containing suitable translation initiation signals preceding an ATG start signal. If desired, methionine at the N-terminus may be cleaved from the mature protein by *in vitro* incubation with cyanogen bromide.

Alternatively, recombinant polyproteins or proteins which are not naturally secreted can be secreted from the insect cell by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in insects. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the translocation of the protein into the endoplasmic reticulum.

After insertion of the DNA sequence and/or the gene encoding the expression product precursor of the protein, an insect cell host is co-transformed with the heterologous DNA of the transfer vector and the genomic DNA of wild type baculovirus -- usually by co-transfection. The promoter and transcription termination sequence of the construct will usually comprise a 2-5kb section of the baculovirus genome. Methods for introducing heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summers and Smith *supra*; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* (1983) 3:2156; and Luckow and Summers (1989)). For example, the insertion can be into a gene such as the polyhedrin gene, by homologous double crossover recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. Miller et al., (1989), *Bioessays* 4:91. The DNA sequence, when cloned in place of the polyhedrin gene in the expression vector, is flanked both 5' and 3' by polyhedrin-specific sequences and is positioned downstream of the polyhedrin promoter.

The newly formed baculovirus expression vector is subsequently packaged into an infectious recombinant baculovirus. Homologous recombination occurs at low frequency (between ~1% and ~5%); thus, the majority of the virus produced after cotransfection is still wild-type virus. Therefore, a method is necessary to identify recombinant viruses. An advantage of the expression system is a visual screen allowing recombinant viruses to be distinguished. The polyhedrin protein, which is produced by the native virus, is produced at very high levels in the nuclei of infected cells at late times after viral infection. Accumulated polyhedrin protein forms occlusion bodies that also contain embedded particles. These occlusion bodies, up to 15 μ m in size, are highly refractile, giving them a bright shiny appearance that is readily visualized under the light microscope. Cells infected with recombinant viruses lack occlusion bodies. To distinguish recombinant virus from wild-type virus, the transfection supernatant is plaqued onto a monolayer of insect cells by techniques known to those skilled in the art. Namely, the plaques are screened under the light microscope for the presence (indicative of wild-type virus)

or absence (indicative of recombinant virus) of occlusion bodies. "Current Protocols in Microbiology" Vol. 2 (Ausubel et al. eds) at 16.8 (Supp. 10, 1990); Summers & Smith, *supra*; Miller et al. (1989).

Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia*: *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni* (WO 89/046699; Carbonell et al., (1985) *J. Virol.* 56:153; Wright (1986) *Nature* 321:718; Smith et al., (1983) *Mol. Cell. Biol.* 3:2156; and see generally, Fraser, et al. (1989) *In Vitro Cell. Dev. Biol.* 25:225).

Cells and cell culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression system; cell culture technology is generally known to those skilled in the art. See, e.g. Summers and Smith *supra*.

The modified insect cells may then be grown in an appropriate nutrient medium, which allows for stable maintenance of the plasmid(s) present in the modified insect host. Where the expression product gene is under inducible control, the host may be grown to high density, and expression induced. Alternatively, where expression is constitutive, the product will be continuously expressed into the medium and the nutrient medium must be continuously circulated, while removing the product of interest and augmenting depleted nutrients. The product may be purified by such techniques as chromatography, e.g. HPLC, affinity chromatography, ion exchange chromatography, etc.; electrophoresis; density gradient centrifugation; solvent extraction, or the like. As appropriate, the product may be further purified, as required, so as to remove substantially any insect proteins which are also secreted in the medium or result from lysis of insect cells, so as to provide a product which is at least substantially free of host debris, e.g. proteins, lipids and polysaccharides.

In order to obtain protein expression, recombinant host cells derived from the transformants are incubated under conditions which allow expression of the recombinant protein encoding sequence. These conditions will vary, dependent upon the host cell selected. However, the conditions are readily ascertainable to those of ordinary skill in the art, based upon what is known in the art.

25 iii. Plant Systems

There are many plant cell culture and whole plant genetic expression systems known in the art. Exemplary plant cellular genetic expression systems include those described in patents, such as: US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, *Phytochemistry* 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the references described above in Vaulcombe et al., *Mol. Gen. Genet.* 209:33-40 (1987); Chandler et al., *Plant Molecular Biology* 3:407-418 (1984); Rogers, *J. Biol. Chem.* 260:3731-3738 (1985); Rothstein et al., *Gene* 55:353-356 (1987); Whittier et al., *Nucleic Acids Research* 15:2515-2535 (1987); Wirsel et al., *Molecular Microbiology* 3:3-14 (1989); Yu et al., *Gene* 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by gibberellic acid can be found in R.L. Jones and J. MacMillin, *Gibberellins*: in: *Advanced Plant Physiology*, Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited, London, pp. 21-52. References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel and Hickey, *Proc. Natl. Acad. Sci.* 84:1337-1339 (1987)

Typically, using techniques known in the art, a desired polynucleotide sequence is inserted into an expression cassette comprising genetic regulatory elements designed for operation in plants. The expression cassette is inserted into a desired expression vector with companion sequences upstream and downstream from the expression cassette suitable for expression in a plant host. The companion sequences will be of plasmid or viral origin and provide necessary characteristics to the vector to permit the vectors to move DNA from an original cloning host, such as bacteria, to the desired plant host. The basic bacterial/plant vector construct will preferably provide a broad host range prokaryote replication origin; a prokaryote selectable marker; and, for Agrobacterium transformations, T DNA sequences for Agrobacterium-mediated transfer to plant chromosomes. Where the heterologous gene is not readily amenable to detection, the construct will preferably also have a selectable marker gene suitable for determining if a plant cell has been transformed. A general review of suitable markers, for example for the members of the grass family, is found in Wilmink and Dons, 1993, *Plant Mol. Biol. Repr.*, 11(2):165-185.

Sequences suitable for permitting integration of the heterologous sequence into the plant genome are also recommended. These might include transposon sequences and the like for homologous recombination as well as Ti sequences which permit random insertion of a heterologous expression cassette into a plant genome. Suitable prokaryote selectable markers include resistance toward antibiotics such as ampicillin or tetracycline. Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art.

The nucleic acid molecules of the subject invention may be included into an expression cassette for expression of the protein(s) of interest. Usually, there will be only one expression cassette, although two or more are feasible. The recombinant expression cassette will contain in addition to the heterologous protein encoding sequence the following elements, a promoter region, plant 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes equipped with one, and a transcription and translation termination sequence. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

A heterologous coding sequence may be for any protein relating to the present invention. The sequence encoding the protein of interest will encode a signal peptide which allows processing and translocation of the protein, as appropriate, and will usually lack any sequence which might result in the binding of the desired protein of the invention to a membrane. Since, for the most part, the transcriptional initiation region will be for a gene which is expressed and translocated during germination, by employing the signal peptide which provides for translocation, one may also provide for translocation of the protein of interest. In this way, the protein(s) of interest will be translocated from the cells in which they are expressed and may be efficiently harvested. Typically secretion in seeds are across the aleurone or scutellar epithelium layer into the endosperm of the seed. While it is not required that the protein be secreted from the cells in which the protein is produced, this facilitates the isolation and purification of the recombinant protein.

Since the ultimate expression of the desired gene product will be in a eucaryotic cell it is desirable to determine whether any portion of the cloned gene contains sequences which will be processed out as introns by the host's splicosome machinery. If so, site-directed mutagenesis of the "intron" region may be conducted to prevent losing a portion of the genetic message as a false intron code, Reed and Maniatis, *Cell* 41:95-105, 1985.

The vector can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. Crossway, *Mol. Gen. Genet.*, 202:179-185, 1985. The genetic material may also be

transferred into the plant cell by using polyethylene glycol, Krens, et al., *Nature*, 296, 72-74, 1982. Another method of introduction of nucleic acid segments is high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface, Klein, et al., *Nature*, 327, 70-73, 1987 and Knudsen and Muller, 1991, *Planta*, 185:330-336 teaching particle bombardment of barley endosperm to create transgenic barley. Yet another method of introduction would be fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies, Fraley, et al., *Proc. Natl. Acad. Sci. USA*, 79, 1859-1863, 1982.

The vector may also be introduced into the plant cells by electroporation. (Fromm et al., *Proc. Natl Acad. Sci. USA* 82:5824, 1985). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersicon*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Hererocallis*, *Nemesia*, *Pelargonium*, *Panicum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browalia*, *Glycine*, *Lolium*, *Zea*, *Triticum*, *Sorghum*, and *Datura*.

Means for regeneration vary from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted. Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

In some plant cell culture systems, the desired protein of the invention may be excreted or alternatively, the protein may be extracted from the whole plant. Where the desired protein of the invention is secreted into the medium, it may be collected. Alternatively, the embryos and embryoless-half seeds or other plant tissue may be mechanically disrupted to release any secreted protein between cells and tissues. The mixture may be suspended in a buffer solution to retrieve soluble proteins. Conventional protein isolation and purification methods will be then used to purify the recombinant protein. Parameters of time, temperature pH, oxygen, and volumes will be adjusted through routine methods to optimize expression and recovery of heterologous protein.

iv. Bacterial Systems

Bacterial expression techniques are known in the art. A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (e.g. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in *Escherichia coli* (*E. coli*) [Raibaud *et al.* (1984) *Annu. Rev. Genet.* 18:173]. Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (*lac*) [Chang *et al.* (1977) *Nature* 198:1056], and maltose. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan (*trp*) [Goeddel *et al.* (1980) *Nuc. Acids Res.* 8:4057; Yelverton *et al.* (1981) *Nucl. Acids Res.* 9:731; US patent 4,738,921; EP-A-0036776 and EP-A-0121775]. The g-lactamase (*bla*) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In *Interferon* 3 (ed. I. Gresser)], bacteriophage lambda PL [Shimatake *et al.* (1981) *Nature* 292:128] and T5 [US patent 4,689,406] promoter systems also provide useful promoter sequences.

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [US patent 4,551,433]. For example, the *tac* promoter is a hybrid *trp-lac* promoter comprised of both *trp* promoter and *lac* operon sequences that is regulated by the *lac* repressor [Amann *et al.* (1983) *Gene* 25:167; de Boer *et al.* (1983) *Proc. Natl. Acad. Sci.* 80:21]. Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA polymerase/promoter system is an example of a coupled promoter system [Studier *et al.* (1986) *J. Mol. Biol.* 189:113; Tabor *et al.* (1985) *Proc. Natl. Acad. Sci.* 82:1074]. In addition, a hybrid promoter can also be comprised of a bacteriophage promoter and an *E. coli* operator region (EPO-A-0 267 851).

In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of foreign genes in prokaryotes. In *E. coli*, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine *et al.* (1975) *Nature* 254:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' end of *E. coli* 16S rRNA [Steitz *et al.* (1979) "Genetic signals and nucleotide sequences in messenger RNA." In *Biological*

Regulation and Development: Gene Expression (ed. R.F. Goldberger)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-binding site [Sambrook *et al.* (1989) "Expression of cloned genes in Escherichia coli." In *Molecular Cloning: A Laboratory Manual*].

A DNA molecule may be expressed intracellularly. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide or by either *in vivo* or *in vitro* incubation with a bacterial methionine N-terminal peptidase (EPO-A-0 219 237).

Fusion proteins provide an alternative to direct expression. Usually, a DNA sequence encoding the N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of a foreign gene and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the foreign gene [Nagai *et al.* (1984) *Nature* 309:810]. Fusion proteins can also be made with sequences from the *lacZ* [Jia *et al.* (1987) *Gene* 60:197], *trpE* [Allen *et al.* (1987) *J. Biotechnol.* 5:93; Makoff *et al.* (1989) *J. Gen. Microbiol.* 135:11], and *Chey* [EP-A-0 324 647] genes. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (e.g. ubiquitin specific processing-protease) to cleave the ubiquitin from the foreign protein. Through this method, native foreign protein can be isolated [Miller *et al.* (1989) *Bio/Technology* 7:698].

Alternatively, foreign proteins can also be secreted from the cell by creating chimeric DNA molecules that encode a fusion protein comprised of a signal peptide sequence fragment that provides for secretion of the foreign protein in bacteria [US patent 4,336,336]. The signal sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

DNA encoding suitable signal sequences can be derived from genes for secreted bacterial proteins, such as the *E. coli* outer membrane protein gene (*ompA*) [Masui *et al.* (1983), in: *Experimental Manipulation of Gene Expression*; Ghrayeb *et al.* (1984) *EMBO J.* 3:2437] and the *E. coli* alkaline phosphatase signal sequence (*phoA*) [Oka *et al.* (1985) *Proc. Natl. Acad. Sci.* 82:7212]. As an additional example, the signal sequence of the alpha-amylase gene from various *Bacillus* strains can be used to secrete heterologous proteins from *B. subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 244 042].

Usually, transcription termination sequences recognized by bacteria are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Transcription termination sequences frequently include DNA sequences of about 50 nucleotides capable of forming stem loop structures that aid in terminating transcription. Examples include transcription termination sequences derived from genes with strong promoters, such as the *trp* gene in *E. coli* as well as other biosynthetic genes.

Usually, the above described components, comprising a promoter, signal sequence (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (e.g. plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a prokaryotic host either for expression or for cloning and amplification. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably contain at least about 10, and more preferably at least about 20 plasmids. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host.

Alternatively, the expression constructs can be integrated into the bacterial genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to the bacterial chromosome that allows the vector to integrate. Integrations appear to result from recombinations between homologous DNA in the vector and the bacterial chromosome. For example, integrating vectors constructed with DNA from various Bacillus strains integrate into the Bacillus chromosome (EP-A-0 127 328). Integrating vectors may also be comprised of bacteriophage or transposon sequences.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of bacterial strains that have been transformed. Selectable markers can be expressed in the bacterial host and may include genes which render bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin (neomycin), and tetracycline [Davies *et al.* (1978) *Annu. Rev. Microbiol.* 32:469]. Selectable markers may also include biosynthetic genes, such as those in the histidine, tryptophan, and leucine biosynthetic pathways.

Alternatively, some of the above described components can be put together in transformation vectors. Transformation vectors are usually comprised of a selectable market that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, *inter alia*, the following bacteria: *Bacillus subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541], *Escherichia coli* [Shimatake *et al.* (1981) *Nature* 292:128; Amann *et al.* (1985) *Gene* 40:183; Studier *et al.* (1986) *J. Mol. Biol.* 189:113; EP-A-0 036 776, EP-A-0 136 829 and EP-A-0 136 907], *Streptococcus cremoris* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655]; *Streptococcus lividans* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655], *Streptomyces lividans* [US patent 4,745,056].

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with CaCl_2 or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. See e.g. [Masson *et al.* (1989) *FEMS Microbiol. Lett.* 60:273; Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, *Bacillus*], [Miller *et al.* (1988) *Proc. Natl. Acad. Sci.* 85:856; Wang *et al.* (1990) *J. Bacteriol.* 172:949, *Campylobacter*], [Cohen *et al.* (1973) *Proc. Natl. Acad. Sci.* 69:2110; Dower *et al.* (1988) *Nucleic Acids Res.* 16:6127; Kushner (1978) "An improved method for transformation of *Escherichia coli* with *ColE1*-derived plasmids. In *Genetic*

Engineering: Proceedings of the International Symposium on Genetic Engineering (eds. H.W. Boyer and S. Nicosia); Mandel *et al.* (1970) *J. Mol. Biol.* 53:159; Taketo (1988) *Biochim. Biophys. Acta* 949:318; Escherichia], [Chassy *et al.* (1987) *FEMS Microbiol. Lett.* 44:173 Lactobacillus]; [Fiedler *et al.* (1988) *Anal. Biochem* 170:38, Pseudomonas]; [Augustin *et al.* (1990) *FEMS Microbiol. Lett.* 66:203, *Staphylococcus*],
5 [Barany *et al.* (1980) *J. Bacteriol.* 144:698; Harlander (1987) "Transformation of *Streptococcus lactis* by electroporation, in: *Streptococcal Genetics* (ed. J. Ferretti and R. Curtiss III); Perry *et al.* (1981) *Infect. Immun.* 32:1295; Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655; Somkuti *et al.* (1987) *Proc. 4th Evr. Cong. Biotechnology* 1:412, *Streptococcus*].

v. Yeast Expression

10 Yeast expression systems are also known to one of ordinary skill in the art. A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (e.g. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the "TATA Box") and a transcription initiation site. A yeast promoter may
15 also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible) expression. Constitutive expression occurs in the absence of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.

20 Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples include alcohol dehydrogenase (ADH) (EP-A-0 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO-A-0 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences [Myanohara *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:1].

25 In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (US Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters which consist of the regulatory sequences of either the *ADH2*,
30 *GAL4*, *GAL10*, OR *PHO5* genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 164 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, *inter alia*, [Cohen *et al.* (1980) *Proc. Natl. Acad. Sci. USA* 77:1078; Henikoff *et al.* (1981) *Nature* 283:835; Hollenberg *et al.* (1981) *Curr. Topics Microbiol. Immunol.* 96:119;
35 Hollenberg *et al.* (1979) "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast *Saccharomyces cerevisiae*," in: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon *et al.* (1980) *Gene* 11:163; Panthier *et al.* (1980) *Curr. Genet.* 2:109;].

A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always

be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast 5 protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See e.g. EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin 10 region that preferably retains a site for a processing enzyme (e.g. ubiquitin-specific processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (e.g. WO88/024066).

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric 15 DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provide for secretion in yeast of the foreign protein. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell.

DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the 20 genes for invertase (EP-A-0012873; JPO 62,096,086) and A-factor (US patent 4,588,684). Alternatively, leaders of non-yeast origin exist, such as an interferon leader, that also provide for secretion in yeast (EP-A-0060057).

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which 25 contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated alpha-factor leaders (usually about 25 to about 50 amino acid residues) (US Patents 4,546,083 and 4,870,008; EP-A-0 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alphafactor. (e.g. see WO 89/02463.)

Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the 30 translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression 35 constructs are often maintained in a replicon, such as an extrachromosomal element (e.g. plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle vectors include YEp24 [Botstein *et al.* (1979) *Gene* 8:17-24], pCI1 [Brake *et al.* (1984) *Proc. Natl. Acad. Sci USA* 81:4642-4646], and YRp17 [Stinchcomb *et al.* (1982) *J. Mol. Biol.* 158:157]. In addition, a replicon may be either a high or low copy number plasmid. A high copy 40

number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Enter a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See e.g. Brake *et al.*, *supra*.

- 5 Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome [Orr-Weaver *et al.* (1983) *Methods in Enzymol.* 101:228-245]. An integrating vector may be
10 directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver *et al.*, *supra*. One or more expression construct may integrate, possibly affecting levels of recombinant protein produced [Rine *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:6750]. The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the chromosome and flanking the
15 expression construct in the vector, which can result in the stable integration of only the expression construct.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers may include biosynthetic genes that can be expressed in the yeast host, such as *ADE2*, *HIS4*, *LEU2*, *TRP1*, and *ALG7*, and the G418 resistance gene, which confer resistance in yeast cells to tunicamycin and G418, respectively. In addition, a suitable selectable marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of *CUP1* allows yeast to grow in the presence of copper ions [Butt *et al.* (1987) *Microbiol. Rev.* 51:351].

20 Alternatively, some of the above described components can be put together into transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or developed into an integrating vector, as described above.
25

Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, *inter alia*, the following yeasts: *Candida albicans* [Kurtz, *et al.* (1986) *Mol. Cell. Biol.* 6:142], *Candida maltosa* [Kunze, *et al.* (1985) *J. Basic Microbiol.* 25:141], *Hansenula polymorpha* [Gleeson, *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302], *Kluyveromyces fragilis* [Das, *et al.* (1984) *J. Bacteriol.* 158:1165], *Kluyveromyces lactis* [De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:737; Van den Berg *et al.* (1990) *Bio/Technology* 8:135], *Pichia guillermondii* [Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141], *Pichia pastoris* [Cregg, *et al.* (1985) *Mol. Cell. Biol.* 5:3376; US Patent Nos. 4,837,148 and 4,929,555],
30 *Saccharomyces cerevisiae* [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163], *Schizosaccharomyces pombe* [Beach and Nurse (1981) *Nature* 300:706], and *Yarrowia lipolytica* [Davidow, *et al.* (1985) *Curr. Genet.* 10:380471 Gaillardin, *et al.* (1985) *Curr. Genet.* 10:49].

40 Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. See e.g. [Kurtz *et al.* (1986) *Mol. Cell. Biol.* 6:142; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; *Candida*]; [Gleeson *et al.* (1986) *J. Gen. Microbiol.* 132:3459;

Roggencamp et al. (1986) *Mol. Gen. Genet.* 202:302; Hansenula]; [Das et al. (1984) *J. Bacteriol.* 158:1165; De Louvencourt et al. (1983) *J. Bacteriol.* 154:1165; Van den Berg et al. (1990) *BioTechnology* 8:135; Kluyveromyces]; [Cregg et al. (1985) *Mol. Cell. Biol.* 5:3376; Kunze et al. (1985) *J. Basic Microbiol.* 25:141; US Patents 4,837,148 & 4,929,555; Pichia]; [Hinnen et al. (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito et al. 5 (1983) *J. Bacteriol.* 153:163 Saccharomyces]; [Beach & Nurse (1981) *Nature* 300:706; Schizosaccharomyces]; [Davidow et al. (1985) *Curr. Genet.* 10:39; Gaillardin et al. (1985) *Curr. Genet.* 10:49; Yarrowia].

Pharmaceutical Compositions

Pharmaceutical compositions can comprise polypeptides and/or nucleic acid of the invention. The pharmaceutical compositions will comprise a therapeutically effective amount of either polypeptides, antibodies, 10 or polynucleotides of the claimed invention.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject 15 will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation can be determined by routine experimentation and is within the judgement of the clinician.

For purposes of the present invention, an effective dose will be from about 0.01 mg/kg to 50 mg/kg or 0.05 20 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and 25 which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in the art.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, 30 hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering 35 substances, and the like, may be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier.

Delivery Methods

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

- 5 Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (e.g. see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Vaccines

- 10 Vaccines according to the invention may either be prophylactic (ie. to prevent infection) or therapeutic (ie. to treat disease after infection).

Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with "pharmaceutically acceptable carriers," which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are 15 typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. 20 pylori*, etc. pathogens.

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59TM (WO 90/14837; Chapter 10 in 25 *Vaccine design: the subunit and adjuvant approach*, eds. Powell & Newman, Plenum Press 1995), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalane, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to 30 generate a larger particle size emulsion, and (c) RibiTM adjuvant system (RAS), (Ribi Immunochem, Hamilton, MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphoryl lipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (DetoxTM); (3) saponin adjuvants, such as StimulonTM (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, such as interleukins (e.g. IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, etc.), interferons (e.g. gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc; and (6) other substances that 35 act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59TM are preferred.

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), etc.

The immunogenic compositions (*e.g.* the immunising antigen/immunogen/polypeptide/protein/nucleic acid, pharmaceutically acceptable carrier, and adjuvant) typically will contain diluents, such as water, saline, glycerol, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.

Typically, the immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above under pharmaceutically acceptable carriers.

Immunogenic compositions used as vaccines comprise an immunologically effective amount of the antigenic or immunogenic polypeptides, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated (*e.g.* nonhuman primate, primate, *etc.*), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials.

The immunogenic compositions are conventionally administered parenterally, *e.g.* by injection, either subcutaneously, intramuscularly, or transdermally/transcutaneously (*e.g.* WO98/20734). Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be administered in conjunction with other immunoregulatory agents.

As an alternative to protein-based vaccines, DNA vaccination may be employed [*e.g.* Robinson & Torres (1997) *Seminars in Immunology* 9:271-283; Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648; see later herein].

Gene Delivery Vehicles

Gene therapy vehicles for delivery of constructs including a coding sequence of a therapeutic of the invention, to be delivered to the mammal for expression in the mammal, can be administered either locally or systemically. These constructs can utilize viral or non-viral vector approaches in *in vivo* or *ex vivo* modality. Expression of such coding sequence can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence *in vivo* can be either constitutive or regulated.

The invention includes gene delivery vehicles capable of expressing the contemplated nucleic acid sequences. The gene delivery vehicle is preferably a viral vector and, more preferably, a retroviral, adenoviral, adeno-associated viral (AAV), herpes viral, or alphavirus vector. The viral vector can also be an astrovirus, coronavirus, orthomyxovirus, papovavirus, paramyxovirus, parvovirus, picornavirus, poxvirus, or togavirus viral vector. See generally, Jolly (1994) *Cancer Gene Therapy* 1:51-64; Kimura (1994) *Human Gene Therapy* 5:845-852; Connelly (1995) *Human Gene Therapy* 6:185-193; and Kaplitt (1994) *Nature Genetics* 6:148-153.

Retroviral vectors are well known in the art and we contemplate that any retroviral gene therapy vector is employable in the invention, including B, C and D type retroviruses, xenotropic retroviruses (for example, NZB-X1, NZB-X2 and NZB9-1 (see O'Neill (1985) *J. Virol.* 53:160) polytropic retroviruses e.g. MCF and MCF-MLV (see Kelly (1983) *J. Virol.* 45:291), spumaviruses and lentiviruses. See RNA Tumor Viruses,

5 Second Edition, Cold Spring Harbor Laboratory, 1985.

Portions of the retroviral gene therapy vector may be derived from different retroviruses. For example, retrovector LTRs may be derived from a Murine Sarcoma Virus, a tRNA binding site from a Rous Sarcoma Virus, a packaging signal from a Murine Leukemia Virus, and an origin of second strand synthesis from an Avian Leukosis Virus.

10 These recombinant retroviral vectors may be used to generate transduction competent retroviral vector particles by introducing them into appropriate packaging cell lines (see US patent 5,591,624). Retrovirus vectors can be constructed for site-specific integration into host cell DNA by incorporation of a chimeric integrase enzyme into the retroviral particle (see WO96/37626). It is preferable that the recombinant viral vector is a replication defective recombinant virus.

15 Packaging cell lines suitable for use with the above-described retrovirus vectors are well known in the art, are readily prepared (see WO95/30763 and WO92/05266), and can be used to create producer cell lines (also termed vector cell lines or "VCLs") for the production of recombinant vector particles. Preferably, the packaging cell lines are made from human parent cells (e.g. HT1080 cells) or mink parent cell lines, which eliminates inactivation in human serum.

20 Preferred retroviruses for the construction of retroviral gene therapy vectors include Avian Leukosis Virus, Bovine Leukemia, Virus, Murine Leukemia Virus, Mink-Cell Focus-Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis Virus and Rous Sarcoma Virus. Particularly preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley and Rowe (1976) *J Virol* 19:19-25), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Graffi, Gross (ATCC Nol VR-590), Kirsten, Harvey Sarcoma Virus and Rauscher (ATCC No. 25 VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190). Such retroviruses may be obtained from depositories or collections such as the American Type Culture Collection ("ATCC") in Rockville, Maryland or isolated from known sources using commonly available techniques.

Exemplary known retroviral gene therapy vectors employable in this invention include those described in patent applications GB2200651, EP0415731, EP0345242, EP0334301, WO89/02468; WO89/05349, WO89/09271, 30 WO90/02806, WO90/07936, WO94/03622, WO93/25698, WO93/25234, WO93/11230, WO93/10218, WO91/02805, WO91/02825, WO95/07994, US 5,219,740, US 4,405,712, US 4,861,719, US 4,980,289, US 4,777,127, US 5,591,624. See also Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53 (1993) 83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci* 81:6349; and Miller (1990) 35 *Human Gene Therapy* 1.

Human adenoviral gene therapy vectors are also known in the art and employable in this invention. See, for example, Berkner (1988) *Biotechniques* 6:616 and Rosenfeld (1991) *Science* 252:431, and WO93/07283, WO93/06223, and WO93/07282. Exemplary known adenoviral gene therapy vectors employable in this invention include those described in the above referenced documents and in WO94/12649, WO93/03769, 40 WO93/19191, WO94/28938, WO95/11984, WO95/00655, WO95/27071, WO95/29993, WO95/34671,

WO96/05320, WO94/08026, WO94/11506, WO93/06223, WO94/24299, WO95/14102, WO95/24297, WO95/02697, WO94/28152, WO94/24299, WO95/09241, WO95/25807, WO95/05835, WO94/18922 and WO95/09654. Alternatively, administration of DNA linked to killed adenovirus as described in Curiel (1992) *Hum. Gene Ther.* 3:147-154 may be employed. The gene delivery vehicles of the invention also include 5 adenovirus associated virus (AAV) vectors. Leading and preferred examples of such vectors for use in this invention are the AAV-2 based vectors disclosed in Srivastava, WO93/09239. Most preferred AAV vectors comprise the two AAV inverted terminal repeats in which the native D-sequences are modified by substitution of nucleotides, such that at least 5 native nucleotides and up to 18 native nucleotides, preferably at least 10 native nucleotides up to 18 native nucleotides, most preferably 10 native nucleotides are retained and the 10 remaining nucleotides of the D-sequence are deleted or replaced with non-native nucleotides. The native D-sequences of the AAV inverted terminal repeats are sequences of 20 consecutive nucleotides in each AAV inverted terminal repeat (ie. there is one sequence at each end) which are not involved in HP formation. The non-native replacement nucleotide may be any nucleotide other than the nucleotide found in the native D-sequence in the same position. Other employable exemplary AAV vectors are pWP-19, pWN-1, both of 15 which are disclosed in Nahreini (1993) *Gene* 124:257-262. Another example of such an AAV vector is psub201 (see Samulski (1987) *J. Virol.* 61:3096). Another exemplary AAV vector is the Double-D ITR vector. Construction of the Double-D ITR vector is disclosed in US Patent 5,478,745. Still other vectors are those disclosed in Carter US Patent 4,797,368 and Muzyczka US Patent 5,139,941, Chartejee US Patent 5,474,935, and Kotin WO94/288157. Yet a further example of an AAV vector employable in this invention is 20 SSV9AFABTKneo, which contains the AFP enhancer and albumin promoter and directs expression predominantly in the liver. Its structure and construction are disclosed in Su (1996) *Human Gene Therapy* 7:463-470. Additional AAV gene therapy vectors are described in US 5,354,678, US 5,173,414, US 5,139,941, and US 5,252,479.

The gene therapy vectors of the invention also include herpes vectors. Leading and preferred examples are 25 herpes simplex virus vectors containing a sequence encoding a thymidine kinase polypeptide such as those disclosed in US 5,288,641 and EP0176170 (Roizman). Additional exemplary herpes simplex virus vectors include HFEM/ICP6-LacZ disclosed in WO95/04139 (Wistar), pHHSVlac described in Geller (1988) *Science* 241:1667-1669 and in WO90/09441 & WO92/07945, HSV Us3::pgC-lacZ described in Fink (1992) *Human Gene Therapy* 3:11-19 and HSV 7134, 2 RH 105 and GAL4 described in EP 0453242 (Breakefield), and those 30 deposited with ATCC as accession numbers ATCC VR-977 and ATCC VR-260.

Also contemplated are alpha virus gene therapy vectors that can be employed in this invention. Preferred alpha virus vectors are Sindbis viruses vectors. Togaviruses, Semliki Forest virus (ATCC VR-67; ATCC VR-1247), Middleberg virus (ATCC VR-370), Ross River virus (ATCC VR-373; ATCC VR-1246), Venezuelan equine encephalitis virus (ATCC VR923; ATCC VR-1250; ATCC VR-1249; ATCC VR-532), and those described in 35 US patents 5,091,309, 5,217,879, and WO92/10578. More particularly, those alpha virus vectors described in US Serial No. 08/405,627, filed March 15, 1995, WO94/21792, WO92/10578, WO95/07994, US 5,091,309 and US 5,217,879 are employable. Such alpha viruses may be obtained from depositories or collections such as the ATCC in Rockville, Maryland or isolated from known sources using commonly available techniques. Preferably, alphavirus vectors with reduced cytotoxicity are used (see USSN 08/679640).

40 DNA vector systems such as eukaryotic layered expression systems are also useful for expressing the nucleic acids of the invention. See WO95/07994 for a detailed description of eukaryotic layered expression systems.

Preferably, the eukaryotic layered expression systems of the invention are derived from alphavirus vectors and most preferably from Sindbis viral vectors.

Other viral vectors suitable for use in the present invention include those derived from poliovirus, for example ATCC VR-58 and those described in Evans, *Nature* 339 (1989) 385 and Sabin (1973) *J. Biol. Standardization*

- 5 1:115; rhinovirus, for example ATCC VR-1110 and those described in Arnold (1990) *J Cell Biochem* L401; pox viruses such as canary pox virus or vaccinia virus, for example ATCC VR-111 and ATCC VR-2010 and those described in Fisher-Hoch (1989) *Proc Natl Acad Sci* 86:317; Flexner (1989) *Ann NY Acad Sci* 569:86, Flexner (1990) *Vaccine* 8:17; in US 4,603,112 and US 4,769,330 and WO89/01973; SV40 virus, for example ATCC VR-305 and those described in Mulligan (1979) *Nature* 277:108 and Madzak (1992) *J Gen Virol* 73:1533;
- 10 influenza virus, for example ATCC VR-797 and recombinant influenza viruses made employing reverse genetics techniques as described in US 5,166,057 and in Enami (1990) *Proc Natl Acad Sci* 87:3802-3805; Enami & Palese (1991) *J Virol* 65:2711-2713 and Luytjes (1989) *Cell* 59:110, (see also McMichael (1983) *NEJ Med* 309:13, and Yap (1978) *Nature* 273:238 and *Nature* (1979) 277:108); human immunodeficiency virus as described in EP-0386882 and in Buchschacher (1992) *J. Virol.* 66:2731; measles virus, for example ATCC VR-67 and VR-1247 and those described in EP-0440219; Aura virus, for example ATCC VR-368; Bebaru virus, for example ATCC VR-600 and ATCC VR-1240; Cabassou virus, for example ATCC VR-922; Chikungunya virus, for example ATCC VR-64 and ATCC VR-1241; Fort Morgan Virus, for example ATCC VR-924; Getah virus, for example ATCC VR-369 and ATCC VR-1243; Kyzylagach virus, for example ATCC VR-927; Mayaro virus, for example ATCC VR-66; Mucambo virus, for example ATCC VR-580 and ATCC VR-1244; Ndumu virus, for example ATCC VR-371; Pixuna virus, for example ATCC VR-372 and ATCC VR-1245; Tonate virus, for example ATCC VR-925; Triniti virus, for example ATCC VR-469; Una virus, for example ATCC VR-374; Whataroa virus, for example ATCC VR-926; Y-62-33 virus, for example ATCC VR-375; O'Nyong virus, Eastern encephalitis virus, for example ATCC VR-65 and ATCC VR-1242; Western encephalitis virus, for example ATCC VR-70, ATCC VR-1251, ATCC VR-622 and ATCC VR-1252; and coronavirus, for example ATCC VR-740 and those described in Hamre (1966) *Proc Soc Exp Biol Med* 121:190.

- 25 Delivery of the compositions of this invention into cells is not limited to the above mentioned viral vectors. Other delivery methods and media may be employed such as, for example, nucleic acid expression vectors, polycationic condensed DNA linked or unlinked to killed adenovirus alone, for example see US Serial No. 08/366,787, filed December 30, 1994 and Curiel (1992) *Hum Gene Ther* 3:147-154 ligand linked DNA, for example see Wu (1989) *J Biol Chem* 264:16985-16987, eucaryotic cell delivery vehicles cells, for example see US Serial No.08/240,030, filed May 9, 1994, and US Serial No. 08/404,796, deposition of photopolymerized hydrogel materials, hand-held gene transfer particle gun, as described in US Patent 5,149,655, ionizing radiation as described in US5,206,152 and in WO92/11033, nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip (1994) *Mol Cell Biol* 14:2411-2418 and in Woffendin (1994) *Proc Natl Acad Sci* 91:1581-1585.

- 35 Particle mediated gene transfer may be employed, for example see US Serial No. 60/023,867. Briefly, the sequence can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, as described in Wu & Wu (1987) *J. Biol. Chem.* 262:4429-4432, insulin as described in Hucked (1990) *Biochem Pharmacol* 40:253-263, galactose as described in Plank (1992) *Bioconjugate Chem* 3:533-539, lactose or transferrin.

Naked DNA may also be employed. Exemplary naked DNA introduction methods are described in WO90/11092 and US 5,580,859. Uptake efficiency may be improved using biodegradable latex beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the beads. The method may be improved further by treatment of the beads to increase hydrophobicity and thereby facilitate disruption of the 5 endosome and release of the DNA into the cytoplasm.

- Liposomes that can act as gene delivery vehicles are described in US 5,422,120, WO95/13796, WO94/23697, WO91/14445 and EP-524,968. As described in USSN. 60/023,867, on non-viral delivery, the nucleic acid sequences encoding a polypeptide can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then be incubated with synthetic gene transfer molecules such as 10 polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, insulin, galactose, lactose, or transferrin. Other delivery systems include the use of liposomes to encapsulate DNA comprising the gene under the control of a variety of tissue-specific or ubiquitously-active promoters. Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al* (1994) *Proc. Natl. Acad. Sci. USA* 91(24):11581-11585. 15 Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun, as described in US 5,149,655; use of ionizing radiation for activating transferred gene, as described in US 5,206,152 and WO92/11033
- 20 Exemplary liposome and polycationic gene delivery vehicles are those described in US 5,422,120 and 4,762,915; in WO 95/13796; WO94/23697; and WO91/14445; in EP-0524968; and in Stryer, Biochemistry, pages 236-240 (1975) W.H. Freeman, San Francisco; Szoka (1980) *Biochem Biophys Acta* 600:1; Bayer (1979) *Biochem Biophys Acta* 550:464; Rivnay (1987) *Meth Enzymol* 149:119; Wang (1987) *Proc Natl Acad Sci* 84:7851; Plant (1989) *Anal Biochem* 176:420.
- 25 A polynucleotide composition can comprises therapeutically effective amount of a gene therapy vehicle, as the term is defined above. For purposes of the present invention, an effective dose will be from about 0.01 mg/kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

Delivery Methods

Once formulated, the polynucleotide compositions of the invention can be administered (1) directly to the 30 subject; (2) delivered *ex vivo*, to cells derived from the subject; or (3) *in vitro* for recombinant protein expression. The subjects to be treated can be mammals or birds. Also, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary 35 administration, suppositories, and transdermal or transcutaneous applications (e.g. see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in e.g. WO93/14778. Examples of cells useful in *ex vivo* applications include, for example, stem cells, particularly hematopoietic, lymph cells, macrophages, dendritic cells, or tumor cells.

Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be accomplished by the following procedures, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

5 Polynucleotide and polypeptide pharmaceutical compositions

In addition to the pharmaceutically acceptable carriers and salts described above, the following additional agents can be used with polynucleotide and/or polypeptide compositions.

A. Polypeptides

One example are polypeptides which include, without limitation: asialoglycosidase (ASOR); transferrin; 10 asialoglycoproteins; antibodies; antibody fragments; ferritin; interleukins; interferons, granulocyte, macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor and erythropoietin. Viral antigens, such as envelope proteins, can also be used. Also, proteins from other invasive organisms, such as the 17 amino acid peptide from the circumsporozoite protein of plasmodium falciparum known as RII.

15 B. Hormones, Vitamins, etc.

Other groups that can be included are, for example: hormones, steroids, androgens, estrogens, thyroid hormone, or vitamins, folic acid.

C. Polyalkylenes, Polysaccharides, etc.

Also, polyalkylene glycol can be included with the desired polynucleotides/polypeptides. In a preferred 20 embodiment, the polyalkylene glycol is polyethylene glycol. In addition, mono-, di-, or polysaccharides can be included. In a preferred embodiment of this aspect, the polysaccharide is dextran or DEAE-dextran. Also, chitosan and poly(lactide-co-glycolide)

D. Lipids, and Liposomes

The desired polynucleotide/polypeptide can also be encapsulated in lipids or packaged in liposomes prior to 25 delivery to the subject or to cells derived therefrom.

Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed polynucleotide to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight (1991) *Biochim. Biophys. Acta.* 1097:1-17; Straubinger (1983) *Meth. Enzymol.* 101:512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations. Cationic liposomes have been shown to mediate intracellular delivery of plasmid DNA (Felgner (1987) *Proc. Natl. Acad. Sci. USA* 84:7413-7416); mRNA (Malone (1989) *Proc. Natl. Acad. Sci. USA* 86:6077-6081); and purified transcription factors (Debs (1990) *J. Biol. Chem.* 35:265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N[1-2,3-dioleyloxy]propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See,

also, Felgner *supra*). Other commercially available liposomes include transfetace (DDAB/DOPE) and DOTAP/DOPE (Boehringer). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, e.g. Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; WO90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphosphatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the 10 DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilamellar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See e.g. Straubinger (1983) *Meth. Immunol.* 101:512-527; Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; Papahadjopoulos (1975) *Biochim. Biophys. Acta* 394:483; Wilson (1979) *Cell* 17:77; Deamer & 15 Bangham (1976) *Biochim. Biophys. Acta* 443:629; Ostro (1977) *Biochem. Biophys. Res. Commun.* 76:836; Fraley (1979) *Proc. Natl. Acad. Sci. USA* 76:3348; Enoch & Strittmatter (1979) *Proc. Natl. Acad. Sci. USA* 76:145; Fraley (1980) *J. Biol. Chem.* (1980) 255:10431; Szoka & Papahadjopoulos (1978) *Proc. Natl. Acad. Sci. USA* 75:145; and Schaefer-Ridder (1982) *Science* 215:166.

20 E.Lipoproteins

In addition, lipoproteins can be included with the polynucleotide/polypeptide to be delivered. Examples of 25 lipoproteins to be utilized include: chylomicrons, HDL, IDL, LDL, and VLDL. Mutants, fragments, or fusions of these proteins can also be used. Also, modifications of naturally occurring lipoproteins can be used, such as acetylated LDL. These lipoproteins can target the delivery of polynucleotides to cells expressing lipoprotein receptors. Preferably, if lipoproteins are including with the polynucleotide to be delivered, no other targeting ligand is included in the composition.

Naturally occurring lipoproteins comprise a lipid and a protein portion. The protein portion are known as apoproteins. At the present, apoproteins A, B, C, D, and E have been isolated and identified. At least two of these contain several proteins, designated by Roman numerals, AI, AII, AIV; CI, CII, CIII.

30 A lipoprotein can comprise more than one apoprotein. For example, naturally occurring chylomicrons comprises of A, B, C, & E, over time these lipoproteins lose A and acquire C and E apoproteins. VLDL comprises A, B, C, & E apoproteins, LDL comprises apoprotein B; HDL comprises apoproteins A, C, & E.

The amino acid of these apoproteins are known and are described in, for example, Breslow (1985) *Annu Rev. Biochem* 54:699; Law (1986) *Adv. Exp Med. Biol.* 151:162; Chen (1986) *J Biol Chem* 261:12918; Kane (1980) 35 *Proc Natl Acad Sci USA* 77:2465; and Utermann (1984) *Hum Genet* 65:232.

Lipoproteins contain a variety of lipids including, triglycerides, cholesterol (free and esters), and phospholipids. The composition of the lipids varies in naturally occurring lipoproteins. For example, chylomicrons comprise mainly triglycerides. A more detailed description of the lipid content of naturally occurring lipoproteins can be found, for example, in *Meth. Enzymol.* 128 (1986). The composition of the lipids are chosen to aid in

conformation of the apoprotein for receptor binding activity. The composition of lipids can also be chosen to facilitate hydrophobic interaction and association with the polynucleotide binding molecule.

Naturally occurring lipoproteins can be isolated from serum by ultracentrifugation, for instance. Such methods are described in *Meth. Enzymol. (supra)*; Pitas (1980) *J. Biochem.* 255:5454-5460 and Mahey (1979) *J Clin.*

5 *Invest* 64:743-750. Lipoproteins can also be produced by *in vitro* or recombinant methods by expression of the apoprotein genes in a desired host cell. See, for example, Atkinson (1986) *Annu Rev Biophys Chem* 15:403 and Radding (1958) *Biochim Biophys Acta* 30: 443. Lipoproteins can also be purchased from commercial suppliers, such as Biomedical Technologies, Inc., Stoughton, Massachusetts, USA. Further description of lipoproteins can be found in Zuckermann *et al.* PCT/US97/14465.

10 **F.Polycationic Agents**

Polycationic agents can be included, with or without lipoprotein, in a composition with the desired polynucleotide/polypeptide to be delivered.

Polycationic agents, typically, exhibit a net positive charge at physiological relevant pH and are capable of neutralizing the electrical charge of nucleic acids to facilitate delivery to a desired location. These agents have

15 both *in vitro*, *ex vivo*, and *in vivo* applications. Polycationic agents can be used to deliver nucleic acids to a living subject either intramuscularly, subcutaneously, etc.

The following are examples of useful polypeptides as polycationic agents: polylysine, polyarginine, polyornithine, and protamine. Other examples include histones, protamines, human serum albumin, DNA

20 binding proteins, non-histone chromosomal proteins, coat proteins from DNA viruses, such as (X174, transcriptional factors also contain domains that bind DNA and therefore may be useful as nucleic acid condensing agents. Briefly, transcriptional factors such as C/CEBP, c-jun, c-fos, AP-1, AP-2, AP-3, CPF, Prot-1, Sp-1, Oct-1, Oct-2, CREP, and TFIID contain basic domains that bind DNA sequences.

Organic polycationic agents include: spermine, spermidine, and putrescine.

The dimensions and of the physical properties of a polycationic agent can be extrapolated from the list above, to construct other polypeptide polycationic agents or to produce synthetic polycationic agents.

Synthetic polycationic agents which are useful include, for example, DEAE-dextran, polybrene. Lipofectin™, and lipofectAMINE™ are monomers that form polycationic complexes when combined with polynucleotides/polypeptides.

Nucleic Acid Hybridisation

30 "Hybridization" refers to the association of two nucleic acid sequences to one another by hydrogen bonding. Typically, one sequence will be fixed to a solid support and the other will be free in solution. Then, the two sequences will be placed in contact with one another under conditions that favor hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase sequence to the solid support (Denhardt's reagent or BLOTO); concentration of the sequences; use of compounds to increase the rate of association of sequences (dextran sulfate or polyethylene glycol); and the stringency of the washing conditions following hybridization. See Sambrook *et al.* [supra] vol.2, chapt.9, pp.9.47 to 9.57.

"Stringency" refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 120 to 200°C below the calculated Tm of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook *et al.* at page 9.50.

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1 μ g for a plasmid or phage digest to 10⁻⁹ to 10⁻⁸ g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of probes can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 μ g of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of 10⁸ cpm/ μ g. For a single-copy mammalian gene a conservative approach would start with 10 μ g of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than 10⁸ cpm/ μ g, resulting in an exposure time of ~24 hours.

Several factors can affect the melting temperature (Tm) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

$$T_m = 81 + 16.6(\log_{10}C_i) + 0.4[\%(G + C)] - 0.6(\%\text{formamide}) - 600/n - 1.5(\%\text{mismatch}).$$

where Ci is the salt concentration (monovalent ions) and n is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) *Anal. Biochem.* 138: 267-284).

In designing a hybridization experiment, some factors affecting nucleic acid hybridization can be conveniently altered. The temperature of the hybridization and washes and the salt concentration during the washes are the simplest to adjust. As the temperature of the hybridization increases (*ie.* stringency), it becomes less likely for hybridization to occur between strands that are nonhomologous, and as a result, background decreases. If the radiolabeled probe is not completely homologous with the immobilized fragment (as is frequently the case in gene family and interspecies hybridization experiments), the hybridization temperature must be reduced, and background will increase. The temperature of the washes affects the intensity of the hybridizing band and the degree of background in a similar manner. The stringency of the washes is also increased with decreasing salt concentrations.

In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology, and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to start with both hybridization and wash conditions which are nonstringent. If non-specific bands or high background are observed after autoradiography, the filter can be washed at high stringency and

reexposed. If the time required for exposure makes this approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

Nucleic Acid Probe Assays

Methods such as PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes according to the invention can determine the presence of cDNA or mRNA. A probe is said to "hybridize" with a sequence of the invention if it can form a duplex or double stranded complex, which is stable enough to be detected.

The nucleic acid probes will hybridize to the Chlamydial nucleotide sequences of the invention (including both sense and antisense strands). Though many different nucleotide sequences will encode the amino acid sequence, the native Chlamydial sequence is preferred because it is the actual sequence present in cells. mRNA represents a coding sequence and so a probe should be complementary to the coding sequence; single-stranded cDNA is complementary to mRNA, and so a cDNA probe should be complementary to the non-coding sequence.

The probe sequence need not be identical to the Chlamydial sequence (or its complement) — some variation in the sequence and length can lead to increased assay sensitivity if the nucleic acid probe can form a duplex with target nucleotides, which can be detected. Also, the nucleic acid probe can include additional nucleotides to stabilize the formed duplex. Additional Chlamydial sequence may also be helpful as a label to detect the formed duplex. For example, a non-complementary nucleotide sequence may be attached to the 5' end of the probe, with the remainder of the probe sequence being complementary to a Chlamydial sequence. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the a Chlamydial sequence in order to hybridize therewith and thereby form a duplex which can be detected.

The exact length and sequence of the probe will depend on the hybridization conditions, such as temperature, salt condition and the like. For example, for diagnostic applications, depending on the complexity of the analyte sequence, the nucleic acid probe typically contains at least 10-20 nucleotides, preferably 15-25, and more preferably ≥ 30 nucleotides, although it may be shorter than this. Short primers generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

Probes may be produced by synthetic procedures, such as the triester method of Matteucci *et al.* [J. Am. Chem. Soc. (1981) 103:3185], or according to Urdea *et al.* [Proc. Natl. Acad. Sci. USA (1983) 80: 7461], or using commercially available automated oligonucleotide synthesizers.

The chemical nature of the probe can be selected according to preference. For certain applications, DNA or RNA are appropriate. For other applications, modifications may be incorporated e.g. backbone modifications, such as phosphorothioates or methylphosphonates, can be used to increase *in vivo* half-life, alter RNA affinity, increase nuclease resistance etc. [e.g. see Agrawal & Iyer (1995) Curr Opin Biotechnol 6:12-19; Agrawal (1996) TIBTECH 14:376-387]; analogues such as peptide nucleic acids may also be used [e.g. see Corey (1997) TIBTECH 15:224-229; Buchardt *et al.* (1993) TIBTECH 11:384-386].

Alternatively, the polymerase chain reaction (PCR) is another well-known means for detecting small amounts of target nucleic acids. The assay is described in: Mullis *et al.* [Meth. Enzymol. (1987) 155: 335-350]; US patents 4,683,195 & 4,683,202. Two 'primers' hybridize with the target nucleic acids and are used to prime the reaction. The primers can comprise sequence that does not hybridize to the sequence of the amplification target (or its

complement) to aid with duplex stability or, for example, to incorporate a convenient restriction site. Typically, such sequence will flank the desired Chlamydial sequence.

A thermostable polymerase creates copies of target nucleic acids from the primers using the original target nucleic acids as a template. After a threshold amount of target nucleic acids are generated by the polymerase,

5 they can be detected by more traditional methods, such as Southern blots. When using the Southern blot method, the labelled probe will hybridize to the Chlamydial sequence (or its complement).

Also, mRNA or cDNA can be detected by traditional blotting techniques described in Sambrook *et al* [supra]. mRNA, or cDNA generated from mRNA using a polymerase enzyme, can be purified and separated using gel electrophoresis. The nucleic acids on the gel are then blotted onto a solid support, such as nitrocellulose. The

10 solid support is exposed to a labelled probe and then washed to remove any unhybridized probe. Next, the duplexes containing the labeled probe are detected. Typically, the probe is labelled with a radioactive moiety.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1-189 show data pertaining to examples 1-189.

Figure 190 shows a representative 2D gel of proteins in elementary bodies.

15 Figure 191 shows an alignment of sequences in five (six) proteins of the invention.

EXAMPLES

The examples indicate *C.pneumoniae* proteins, together with evidence to support the view that the proteins are useful antigens for vaccine production and development or for diagnostic purposes. This evidence takes the form of:

- 20 • Computer prediction based on sequence information from CWL029 strain (*e.g.* using the PSORT algorithm available from www.psort.nibb.ac.jp).
- Data on recombinant expression and purification of the proteins cloned from IOL207 strain.
- Western blots to demonstrate immunoreactivity in serum (typically a blot of an EB extract of *C.pneumoniae* strain FB/96 stained with mouse antiserum against the recombinant protein).
- 25 • FACS analysis of *C.pneumoniae* bacteria or purified EBs to confirm accessibility of the antigen to the immune system (see also table III).
- An indication if the protein was identified by MALDI-TOF from a 2D gel electrophoresis map of proteins from purified elementary bodies from strain FB/96. This confirms that the protein is expressed *in vivo* (see also table V).
- 30 Various tests can be used to assess the *in vivo* immunogenicity of the proteins identified in the examples. For example, the proteins can be expressed recombinantly and used to screen patient sera by immunoblot. A positive reaction between the protein and patient serum indicates that the patient has previously mounted an immune response to the protein in question *i.e.* the protein is an immunogen. This method can also be used to identify immunodominant proteins.

The recombinant protein can also be conveniently used to prepare antibodies e.g. in a mouse. These can be used for direct confirmation that a protein is located on the cell-surface. Labelled antibody (e.g. fluorescent labelling for FACS) can be incubated with intact bacteria and the presence of label on the bacterial surface confirms the location of the protein.

- 5 In particular, the following methods (A) to (O) were used to express, purify and biochemically characterise the proteins of the invention:

CLONING OF CPN ORFs FOR EXPRESSION IN *E.COLI*

ORFs of *Chlamydia pneumoniae* (Cpn) were cloned in such a way as to potentially obtain three different kind of proteins:

- 10 a) proteins having an hexa-histidine tag at the C-terminus (cpn-His)
 b) proteins having a GST fusion partner at the N-terminus (Gst-cpn)
 c) proteins having both hexa-histidine tag at the C-terminus and GST at the N-terminus (GST/His fusion; NH₂-GST-cpn-(His)₆-COOH)

The type a) proteins were obtained upon cloning in the pET21b+ (Novagen). The type b) and c) 15 proteins were obtained upon cloning in modified pGEX-KG vectors [Guan & Dixon (1991) *Anal. Biochem.* 192:262]. For instance pGEX-KG was modified to obtain pGEX-NN, then by modifying pGEX-NN to obtain pGEX-NNH. The Gst-cpn and Gst-cpn-His proteins were obtained in pGEX-NN and pGEX-NNH respectively.

The modified versions of pGEX-KG vector were made with the aim of allowing the cloning of 20 single amplification products in all three vectors after only one double restriction enzyme digestion and to minimise the presence of extraneous amino acids in the final recombinant proteins.

(A) Construction of pGEX-NN and pGEX-NNH expression vectors

Two couples of complementary oligodeoxyribonucleotides were synthesised using the DNA 25 synthesiser ABI394 (Perkin Elmer) and the reagents from Cruachem (Glasgow, Scotland). Equimolar amounts of the oligo pairs (50 ng each oligo) were annealed in T4 DNA ligase buffer (New England Biolabs) for 10 min in a final volume of 50µl and then were left to cool slowly at room temperature. With the described procedure he following DNA linkers were obtained:

gexNN linker:

30 NdeI NheI XmaI EcoRI NcoI SalI XbaI SacI NotI
 GATCCC ATATGGCTAGCCGGGGAAATT CGTCCATGGAGTGACTGACTCGAGTGATCGAGCT CCTGAGCGGCCGATGAA
 GGTATACCGATCGGGCCCTTAAGCAGGTACCTCACTCAGCTGACTGAGCTCACTAGCTCGAGGACTCGCCGGTACTTCGA

gexNNH linker:

35 HindIII NotI XbaI --Hexa-Histidine--
 TCGACAAGCTTGC GGCGCACTCGAGCATCACC ATCACC ATCACTGAT
 GTTCGAACGCCGGCTGAGCACGTAGAGGTAGTGGTAGT GACTATCGA

The plasmid pGEX-KG was digested with BamHI and HindIII and 100 ng were ligated overnight at 16 °C to the linker gexNN with a molar ratio of 3:1 linker/plasmid using 200 units of T4 DNA ligase

(New england Biolabs). After transformation of the ligation product in *E. coli* DH5, a clone containing the pGEX-NN plasmid, having the correct linker, was selected by means of restriction enzyme analysis and DNA sequencing.

The new plasmid pGEX-NN was digested with SalI and HindIII and ligated to the linker gexNNH.

- 5 After transformation of the ligation product in *E. coli* DH5, a clone containing the pGEX-NNH plasmid, having the correct linker, was selected by means of restriction enzyme analysis and DNA sequencing.

(B) Chromosomal DNA preparation

The chromosomal DNA of elementary bodies (EB) of *C.pneumoniae* strain 10L-207 was prepared by

- 10 adding 1.5 ml of lysis buffer (10 mM Tris-HCl, 150 mM NaCl, 2 mM EDTA, 0,6 % SDS, 100 µg/ml Proteinase K, pH 8) to 450 µl EB suspension (400.000/µl) and incubating overnight at 37 °C. After sequential extraction with phenol, phenol-chloroform, and chloroform, the DNA was precipitated with 0,3 M sodium acetate, pH 5,2 and 2 volumes of absolute ethanol. The DNA pellet was washed with 70 % ethanol. After solubilization with distilled water and treatment with 20 µg/ml RNase A
15 for 1 hour at RT, the DNA was extracted again with phenol-chloroform, alcohol precipitated and suspended with 300 µl 1 mM Tris-HCl pH 8,5. The DNA concentration was evaluated by measuring OD₂₆₀ of the sample.

(C) Oligonucleotide design

Synthetic oligonucleotide primers were designed on the basis of the coding sequence of each ORF

- 20 using the sequence of *C.pneumoniae* strain CWL029. Any predicted signal peptide were omitted, by deducing the 5' end amplification primer sequence immediately downstream from the predicted leader sequence. For most ORFs, the 5' tail of the primers (table I) included only one restriction enzyme recognition site (NdeI, or NheI, or SpeI depending on the gene's own restriction pattern); the 3' primer tails (tableI) included a XhoI or a NotI or a HindIII restriction site.

	5' tails		3' tails
NdeI	5' GTGCGTCATATG 3'	XhoI	5' GCGTCTCGAG 3'
NheI	5' GTGCGTGCTAGC 3'	NotI	5' ACTCGCTAGCGGCCGC 3'
SpeI	5' GTGCGTACTAGT 3'	HindIII	5' GCGTAAGCTT 3'

25 **Table I.** Oligonucleotide tails of the primers used to amplify Cpn genes.

As well as containing the restriction enzyme recognition sequences, the primers included nucleotides which hybridized to the sequence to be amplified. The number of hybridizing nucleotides depended on the melting temperature of the primers which was determined as described [(Breslauer *et al.* (1986) PNAS USA 83:3746-50]. The average melting temperature of the selected oligos was 50-55°C for the hybridizing region alone and 65-75°C for the whole oligos. Table II shows the forward and reverse primers used for each amplification.

(D) Amplification

The standard PCR protocol was as follow: 50 ng genomic DNA were used as template in the presence of 0,2 μ M each primer, 200 μ M each dNTP, 1,5 mM MgCl₂, 1x PCR buffer minus Mg (Gibco-BRL), and 2 units of Taq DNA polymerase (Platinum Taq, Gibco-BRL) in a final volume of

- 5 100 μ l. Each sample underwent a double-step amplification: the first 5 cycles were performed using as the hybridizing temperature the one of the oligos excluding the restriction enzyme tail, followed by 25 cycles performed according to the hybridization temperature of the whole lenght primers. The standard cycles were as follow:

denaturation : 94 °C, 2 min

10

denaturation: 94 °C, 30 seconds
hybridization: 51 °C, 50 seconds } 5 cycles
elongation: 72 °C, 1 min or 2 min and 40 sec

15

denaturation: 94 °C, 30 seconds
hybridization: 70 °C, 50 seconds } 25 cycles
elongation: 72 °C, 1 min or 2 min and 40 sec

72 °C, 7 min

20 4 °C

The elongation time was 1 min for ORFs shorter than 2000 bp, and 2 min and 40 seconds for ORFs longer than 2000 bp. The amplifications were performed using a Gene Amp PCR system 9600 (Perkin Elmer).

- 25 To check the amplification results, 4 μ l of each PCR product was loaded onto 1-1.5 agarose gel and the size of amplified fragments compared with DNA molecular weight standards (DNA markers III or IX, Roche). The PCR products were loaded on agarose gel and after electrophoresis the right size bands were excised from the gel. The DNA was purified from the agarose using the Gel Extraction Kit (Qiagen) following the instruction of the manufacturer. The final elution volume of the DNA was
30 50 μ l TE (10 mM Tris-HCl, 1 mM EDTA, pH 8). One μ l of each purified DNA was loaded onto agarose gel to evaluate the yield.

(E) Digestion of PCR fragments

One-two μ g of purified PCR product were double digested overnight at 37 °C with the appropriate restriction enzymes (60 units of each enzyme) using the appropriate restriction buffer in 100 μ l final volume. The restriction enzymes and the digestion buffers were from New England Biolabs. After

35

purification of the digested DNA (PCR purification Kit, Qiagen) and elution with 30 µl TE, 1 µl was subjected to agarose gel electrophoresis to evaluate the yield in comparison to titrated molecular weight standards (DNA markers III or IX, Roche).

(F) Digestion of the cloning vectors (pET21b+, pGEX-NN, and pGEX-NNH)

5 10 µg of plasmid was double digested with 100 units of each restriction enzyme in 400 µl reaction volume in the presence of appropriate buffer by overnight incubation at 37 °C. After electrophoresis on a 1% agarose gel, the band corresponding to the digested vector was purified from the gel using the Qiagen Qiaex II Gel Extraction Kit and the DNA was eluted with 50 µl TE. The DNA concentration was evaluated by measuring OD₂₆₀ of the sample.

10 **(G) Cloning**

75ng of the appropriately digested and purified vectors and the digested and purified fragments corresponding to each ORF, were ligated in final volumes of 10-20 µl with a molar ratio of 1:1 fragment/vector, using 400 units T4 DNA ligase (New England Biolabs) in the presence of the buffer supplied by the manufacturer. The reactions were incubated overnight at 16 °C.

15 Transformation in *E. coli* DH5 competent cells was performed as follow: the ligation reaction was mixed with 200 µl of competent DH5 cells and incubated on ice for 30 min and then at 42 °C for 90 seconds. After cooling on ice, 0.8 ml LB was added and the cells were incubated for 45 min at 37 °C under shaking. 100 and 900 µl of cell suspensions were plated on separate plates of agar LB 100 µg/ml Ampicillin and the plates were incubated overnight at 37 °C. The screening of the 20 transformants was done by growing randomly chosen clones in 6 ml LB 100 µg/ml Ampicillin, by extracting the DNA using the Qiagen Qiaprep Spin Miniprep Kit following the manufacturer instructions, and by digesting 2 µl of plasmid minipreparation with the restriction enzymes specific for the restriction cloning sites. After agarose gel electrophoresis of the digested plasmid mini-preparations, positive clones were chosen on the basis of the correct size of the restriction fragments, 25 as evaluated by comparison with appropriate molecular weight markers (DNA markers III or IX, Roche).

(H) Expression

1 µl of each right plasmid mini-preparation was transformed in 200 µl of competent *E. coli* strain suitable for expression of the recombinant protein. All pET21b+ recombinant plasmids were 30 transformed in BL21 DE3 (Novagen) *E. coli* cells, whilst all pGEX-NN and all pGEX-NNH recombinant plasmids were transformed in BL21 cells (Novagen). After plating transformation mixtures on LB/Amp agar plates and incubation overnight at 37 °C, single colonies were inoculated in 3 ml LB 100 µg/ml Ampicillin and grown at 37 °C overnight. 70 µl of the overnight culture was inoculated in 2 ml LB/Amp and grown at 37 °C until OD₆₀₀ of the pET clones reached the 0,4-0,8 35 value or until OD₆₀₀ of the pGEX clones reached the 0,8-1 value. Protein expression was then

induced by adding IPTG (Isopropil β -D thio-galacto-piranoside) to the mini-cultures. pET clones were induced using 1 mM IPTG, whilst pGEX clones were induced using 0.2 mM IPTG. After 3 hours incubation at 37 °C the final OD₆₀₀ was checked and the cultures were cooled on ice. After centrifugation of 0.5 ml culture, the cell pellet was suspended in 50 μ l of protein Loading Sample Buffer (60 mM TRIS-HCl pH 6.8, 5% w/v SDS, 10% v/v glycerin, 0.1% w/v Bromophenol Blue, 100 mM DTT) and incubated at 100 °C for 5 min. A volume of boiled sample corresponding to 0.1 OD₆₀₀ culture was analysed by SDS-PAGE and Coomassie Blue staining to verify the presence of induced protein band.

PURIFICATION OF THE RECOMBINANT PROTEINS

Single colonies were inoculated in 25 ml LB 100 μ g/ml Ampicillin and grown at 37 °C overnight. The overnight culture was inoculated in 500 ml LB/Amp and grown under shaking at 25 °C until OD₆₀₀ 0,4-0,8 value for the pET clones, or until OD₆₀₀ 0,8-1 value for the pGEX clones. Protein expression was then induced by adding IPTG to the cultures. pET clones were induced using 1 mM IPTG, whilst pGEX clones were induced using 0.2 mM IPTG. After 4 hours incubation at 25 °C the final OD₆₀₀ was checked and the cultures were cooled on ice. After centrifugation at 6000 rpm (JA10 rotor, Beckman), the cell pellet was processed for purification or frozen at -20 °C.

(I) Procedure for the purification of soluble His-tagged proteins from *E.coli*

1. Transfer the pellets from -20°C to ice bath and reconstitute with 10 ml 50 mM NaHPO₄ buffer, 300 mM NaCl, pH 8,0, pass in 40-50 ml centrifugation tubes and break the cells as per the following outline:
2. Break the pellets in the French Press performing three passages with in-line washing.
3. Centrifuge at about 30-40000 x g per 15-20 min. If possible use rotor JA 25.50 (21000 rpm, 15 min.) or JA-20 (18000 rpm, 15 min.)
4. Equilibrate the Poly-Prep columns with 1 ml Fast Flow Chelating Sepharose resin with 50 mM phosphate buffer, 300 mM NaCl, pH 8,0.
5. Store the centrifugation pellet at -20°C, and load the supernatant in the columns.
6. Collect the flow through.
7. Wash the columns with 10 ml (2 ml + 2 ml + 4 ml) 50 mM phosphate buffer, 300 mM NaCl, pH 8,0.
- 30 8. Wash again with 10 ml 20 mM imidazole buffer, 50 mM phosphate, 300 mM NaCl, pH 8,0.
9. Elute the proteins bound to the columns with 4,5 ml (1,5 ml + 1,5 ml + 1,5 ml) 250 mM imidazole buffer, 50 mM phosphate, 300 mM NaCl, pH 8,0 and collect the 3 corresponding fractions of ~1,5 ml each. Add to each tube 15 μ l DTT 200 mM (final concentration 2 mM)

-37-

10. Measure the protein concentration of the first two fractions with the Bradford method, collect a 10 µg aliquot of proteins from each sample and analyse by SDS-PAGE. (N.B.: should the sample be too diluted, load 21 µl + 7 µl loading buffer).
11. Store the collected fractions at +4°C while waiting for the results of the SDS-PAGE analysis.
- 5 12. For immunisation prepare 4-5 aliquots of 100 µg each in 0,5 ml in 40% glycerol. The dilution buffer is the above elution buffer, plus 2 mM DTT. Store the aliquots at -20°C until immunisation.

(J) Purification of His-tagged proteins from Inclusion bodies

Purifications were carried out essentially according the following protocol:

- 10 1. Bacteria are collected from 500 ml cultures by centrifugation. If required store bacterial pellets at -20°C. For extraction, resuspend each bacterial pellet in 10 ml 50 mM TRIS-HCl buffer, pH 8,5 on an ice bath.
2. Disrupt the resuspended bacteria with a French Press, performing two passages.
3. Centrifuge at 35000 x g for 15 min and collect the pellets. Use a Beckman rotor JA 25.50 (21000 rpm, 15 min.) or JA-20 (18000 rpm, 15 min.).
- 15 4. Dissolve the centrifugation pellets with 50 mM TRIS-HCl, 1 mM TCEP {Tris(2-carboxyethyl)-phosphine hydrochloride, Pierce} , 6M guanidium chloride, pH 8,5. Stir for ~ 10 min. with a magnetic bar.
5. Centrifuge as described above, and collect the supernatant..
- 20 6. Prepare an adequate number of Poly-Prep (Bio-Rad) columns containing 1 ml of Fast Flow Chelating Sepharose (Pharmacia) saturated with Nichel according to manufacturer recommendations.. Wash the columns twice with 5 ml of H₂O and equilibrate with 50 mM TRIS-HCl, 1 mM TCEP, 6M guanidinium chloride, pH 8,5.
7. Load the supernatants from step 5 onto the columns, and wash with 5 ml of 50 mM TRIS-Hcl buffer, 1 mM TCEP, 6M urea, pH 8,5
- 25 8. Wash the columns with 10 ml of 20 mM imidazole, 50 mM TRIS-HCl , 6M urea, 1 mM TCEP, pH 8,5. Collect and set aside the first 5 ml for possible further controls.
9. Elute the proteins bound to the columns with 4,5 ml of a buffer containing 250 mM imidazole, 50 mM TRIS-HCl, 6M urea, 1 mM TCEP, pH 8,5. Add the elution buffer in three 1,5 ml aliquots, and collect the corresponding 3 fractions. Add to each fraction 15 µl DTT (final concentration 2 mM) .
- 30 10. Measure eluted protein concentration with the Bradford method, and analyze aliquots of ca 10 µg of protein by SDS-PAGE.
11. Store proteins at -20°C in 40% (v/v) glycerol, 50 mM TRIS-HCl, 2M urea, 0.5 M arginine, 2 mM DTT, 0.3 mM TCEP, 83.3 mM imidazole, pH 8,5
- 35

(K) Procedure for the purification of GST-fusion proteins from *E.coli*

1. Transfer the bacterial pellets from -20°C to an ice bath and resuspend with 7,5 ml PBS, pH 7,4 to which a mixture of protease inhibitors (CØMPLETE™ - Boehringer Mannheim, 1 tablet every 25 ml of buffer) has been added. Transfer to 40-50 ml centrifugation tubes and sonicate according to the following procedure:
 - a) Position the probe at about 0,5 cm from the bottom of the tube
 - b) Block the tube with the clamp
 - c) Dip the tube in an ice bath
 - d) Set the sonicator as follows: Timer → Hold, Duty Cycle → 55, Out. Control → 6.
 - e) perform 5 cycles of 10 impulses at a time lapse of 1 minute (i.e. one cycle = 10 impulses + ~45" hold; b. 10 impulses + ~45" hold; c. 10 impulses + ~45" hold; d. 10 impulses + ~45" hold; e. 10 impulses + ~45" hold)
10. Centrifuge at about 30-40000 x g for 15-20 min. E.g.: use rotor Beckman JA 25.50 at 21000 rpm, for 15 min.
15. Store the centrifugation pellets at -20°C, and load the supernatants on the chromatography columns, as follows
20. Equilibrate the Poly-Prep (Bio-Rad) columns with 0,5 ml (\cong 1 ml suspension) of Glutathione-Sepharose 4B resin, wash with 2 ml (1 + 1) H₂O, and then with 10 ml (2 + 4 + 4) PBS, pH 7,4.
25. Load the supernatants on the columns and discard the flow through.
30. Wash the columns with 10 ml (2 + 4 + 4) PBS, pH 7,4.
- Elute the proteins bound to the columns with 4,5 ml of 50 mM TRIS buffer, 10 mM reduced glutathione, pH 8,0, adding 1,5 ml + 1,5 ml + 1,5 ml and collecting the respective 3 fractions of ~1,5 ml each.
- Measure the protein concentration of the first two fractions with the Bradford method, analyse a 10 µg aliquot of proteins from each sample by SDS-PAGE. (N.B.: if the sample is too diluted load 21 µl (+ 7 µl loading buffer).
- Store the collected fractions at +4°C while waiting for the results of the SDS-PAGE analysis.
- For each protein destined to the immunisation prepare 4-5 aliquots of 100 µg each in 0,5 ml of 40% glycerol. The dilution buffer is 50 mM TRIS.HCl, 2 mM DTT, pH 8,0. Store the aliquots at -20°C until immunisation..

SEROLOGY**(L) Protocol of immunization**

1. Groups of four CD1 female mice aged between 6 and 7 weeks were immunized with 20 µg of recombinant protein resuspended in 100 µl.

2. Four mice for each group received 3 doses with a 14 days interval schedule.
3. Immunization was performed through intra-peritoneal injection of the protein with an equal volume of Complete Freund's Adjuvant (CFA) for the first dose and Incomplete Freund's Adjuvant (IFA) for the following two doses.
5. 4. Sera were collected before each immunization. Mice were sacrificed 14 days after the third immunization and the collected sera were pooled and stored at -20°C.

(M) Western blot analysis of Cpn elementary body proteins with mouse sera

Aliquots of elementary bodies containing approximately 4 µg of proteins, mixed with SDS loading buffer (1x: 60 mM TRIS-HCl pH 6.8, 5% w/v SDS, 10% v/v glycerin, 0.1% Bromophenol Blue, 100 mM DTT) and boiled 5 minutes at 95° C, were loaded on a 12% SDS-PAGE gel. The gel was run using a SDS-PAGE running buffer containing 250 mM TRIS, 2.5 mM Glycine and 0.1 %SDS. The gel was electroblotted onto nitrocellulose membrane at 200 mA for 30 minutes. The membrane was blocked for 30 minutes with PBS, 3% skimmed milk powder and incubated O/N at 4° C with the appropriate dilution (1/100) of the sera. After washing twice with PBS + 0.1% Tween (Sigma) the membrane was incubated for 2 hours with peroxidase-conjugated secondary anti-mouse antibody (Sigma) diluted 1:3000. The nitrocellulose was washed twice for 10 minutes with PBS + 0.1% Tween-20 and once with PBS and thereafter developed by Opti-4CN Substrate Kit (Biorad).

Lanes shown in Western blots are: (P) = pre-immune control serum; (I) = immune serum.

(N) FACS analysis of *Chlamydia pneumoniae* elementary bodies with mouse sera

20. 1. 2×10^5 Elementary Bodies (EB)/well were washed with 200 µl of PBS-0.1%BSA in a 96 wells U bottom plate and centrifuged for 10 min. at 1200rpm, at 4°C.
2. The supernatant was discarded and the E.B. resuspended in 10 µl of PBS-0.1%BSA.
3. 10µl mouse sera diluted in PBS-0.1%BSA were added to the E.B. suspension to a final dilution of 1:400, and incubated on ice for 30 min.
25. 4. EB were washed by adding 180µl PBS-0.1%BSA and centrifuged for 10min. at 1200rpm, 4°C.
5. The supernatant was discarded and the E.B. resuspended in 10 l of PBS-0.1%BSA.
6. 10µl of a goat anti-mouse IgG, F(ab')₂ fragment specific-R-Phycoerythrin-conjugated (Jackson Immunoresearch Laboratories Inc., cat.N°115-116-072) was added to the EB suspension to a final dilution of 1:100, and incubated on ice for 30 min. in the dark.
30. 7. EB were washed by adding 180µl PBS-0.1%BSA and centrifuged for 10min. at 1200rpm, 4°C.
8. The supernatant was discarded and the E.B. resuspended in 150 µl of PBS-0.1%BSA.
9. E.B. suspension was passed through a cytometric chamber of a FACS Calibur (Becton Dickinson, Mountain View, CA USA) and 10.000 events were acquired.

10. Data were analysed using Cell Quest Software (Becton Dickinson, Mountain View, CA USA) by drawing a morphological dot plot (using forward and side scatter parameters) on E.B. signals. An histogram plot was then created on FL2 intensity of fluorescence log scale recalling the morphological region of EB.
- 5 NB: the results of FACS depend not only on the extent of accessibility of the native antigens but also on the quality of the antibodies elicited by the recombinant antigens, which may have structures with a variable degree of correct folding as compared with the native protein structures. Therefore, even if a FACS assay appears negative this does not necessarily mean that the protein is not abundant or accessible on the surface. PorB antigen, for instance, gave negative results in FACS but is a surface-exposed neutralising antigen [Kubo & Stephens (2000) *Mol. Microbiol.* 38:772-780].
- 10

(O) Mass Spectrometry analysis of two-dimensional electrophoretic protein maps

Gradient purified EBs from strain FB/96 were solubilized at a final concentration of 5.5mg/ml with immobiline rehydration buffer (7M urea, 2M thiourea, 2% (w/v) CHAPS, 2% (w/v) ASB 14 [Chevallet *et al.* (1998) *Electrophor.* 19:1901-9], 2% (v/v) C.A 3-10NL (Amersham Pharmacia Biotech), 2 mM tributyl phosphine, 65 mM DTT). Samples (250µg protein) were adsorbed overnight on Immobiline DryStrips (7 cm, pH 3-10 non linear). Electrophocusing was performed in a IPGphor Isoelectric Focusing Unit (Amersham Pharmacia Biotech). Before PAGE separation, the focused strips were incubated in 4M urea, 2M thiourea, 30% (v/v) glycerol, 2% (w/v) SDS, 5mM tributyl phosphine 2.5%(w/v) acrylamide, 50mM Tris-HCl pH 8.8, as described [Herbert *et al.* (1998) *Electrophor.* 19:845-51]. SDS-PAGE was performed on linear 9-16% acrylamide gradients. Gels were stained with colloidal Coomassie (Novex, San Diego) [Doherty *et al.* (1998) *Electrophor.* 19:355-63]. Stained gels were scanned with a Personal Densitometer SI (Molecular Dynamics) at 8 bits and 50µm per pixel. Map images were annotated with the software Image Master 2D Elite, version 3.10 (Amersham Pharmacia Biotech). Protein spots were excised from the gel, using an Ettan 25 Spot picker (Amersham Pharmacia Biotech), and dried in a vacuum centrifuge. In-gel digestion of samples for mass spectrometry and extraction of peptides were performed as described by Wilm *et al.* [*Nature* (1996) 379:466-9]. Samples were desalted with a ZIP TIP (Millipore), eluted with a saturated solution of alpha-cyano-4-hydroxycinnamic acid in 50% acetonitrile, 0.1% TFA and directly loaded onto a SCOUT 381 multiprobe plate (Bruker). Spectra were acquired on a Bruker Biflex II MALDI-TOF. Spectra were calibrated using a combination of known standard peptides, located in spots adjacent to the samples. Resulting values for monoisotopic peaks were used for database searches using the computer program Mascot (www.matrixscience.com). All searches were performed using an error of 200-500ppm as constraint. A representative gel is shown in Figure 190.

30

Example 1

- 35 The following *C.pneumoniae* protein (PID 4376552) was expressed <SEQ ID 1; cp6552>:

1 MKKKLSLLVG LIFVLSSCHK EDAQNKRIV ASPTPHAEELL ESLQEEAKDL

-41-

5 51 GIKLKILPVD DYRIPNRLLL DKQVDANYFQ HQAFLDDECE RYDCKGELVV
 101 101 IAKVHLEPQA IYSKKHSSLE RLKSQKKLTI AIPVDRTNAQ RALHLLEECG
 151 151 LIVCKGPANL NMTAKDVCGK ENRSINILEV SAPLLVGSLP DVDAAVIPGN
 201 201 FAIAANLSPK KDSLCLLEDLS VSKYTNLVVI RSEDVGSPKM IKLQKLFQSP
 251 251 SVQHFFDTKY HGNILTMTQD NG*

5 A predicted signal peptide is highlighted.

The cp6552 nucleotide sequence <SEQ ID 2> is:

10 1 ATGAAAAAAA AATTATCATT ACTTGTAGGT TTAATTTTG TTTTGGATTC
 51 51 TTGCCATAAG GAAGATGCTC AGAATAAAAT ACGTATTGTA GCCAGTCCGA
 101 101 CACCTCATGC GGAATTATTG GAGAGTTAC AGGAAGAGGC TAAAGATCTT
 151 151 GGAATCAAGC TGAAAATACT TCCAGTAGAT GATTATCGTA TTCCTAATCG
 201 201 TTTGCTTTTG GATAAACAAAG TAGATGCAA TTACTTTCAA CATCAAGCTT
 251 251 TTCTTGATGA CGAATGCGAG CGTTATGATT GTAAGGGTGA ATTAGTTGTT
 301 301 ATCGCTAAAG TTCATTGGA ACCTCAAGCA ATTTATTCTA AGAAACATTC
 351 351 TTCTTTAGAG CGCTTAAAAAA GCCAGAAGAA ACTGACTATA GCGATTCTG
 401 401 TGGATCGTAC GAATGCTCAG CGTGTCTAC ACTTGTAGA AGAGTGCAGGA
 451 451 CTCATTTGTT GCAGGAGGCC TGCTAATTAA ATATGACAG CTAAGATGT
 501 501 CTGTGGGAAA GAAAATAGAA GTATCAACAT ATTAGAGGTG TCAGCTCCTC
 551 551 TTCTTGTCGG ATCTCTTCCT GACGTTGATG CTGCTGTCAT TCCTGGAAAT
 601 601 TTTGCTATAG CAGCAAACCT TTCTCCAAAG AAAGATAGTC TTTGTTTAGA
 651 651 GGATCTTTCG GTATCTAAGT ATACAAACCT TGTGTCATT CGTTCTGAAG
 701 701 ACGTAGGTTTC TCCTAAAATG ATAAAATTAC AGAACGCTGTT TCAATCTCCT
 751 751 TCTGTACAAAC ATTTTTGTA TACAAAATAT CATGGAAATA TTTGACAAT
 801 801 GACTCAAGAC AATGGTTAG

25 The PSORT algorithm predicts an inner membrane location (0.127).

The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 1A, and also as a GST-fusion. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 1B) and for FACS analysis (Figure 1C).

The cp6552 protein was also identified in the 2D-PAGE experiment (Cpn0278).

30 These experiments show that cp6552 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 2

The following *C.pneumoniae* protein (PID 4376736) was expressed <SEQ ID 3; cp6736>:

35 1 MKTTSIRKFLI STTLAPCFAS TAFTVEVIMP SENFDGSSGK IFPYTTLSDP
 51 51 RGTCIFSGD LYIANLDNAI SRTSSSCFSN RAGALQILGK GGVFSFLNIR
 101 101 SSADGAAISS VITQNPELCP LSFSGFSQMI FDNCESLTSD TSASNVIPHA
 151 151 SAIYATTPML FTNNDSILFQ YNRSAGFGAA IRGTSITIEN TKKSSLFNGN
 201 201 GSISNGGALT GSAAINLINN SAPVIFSTNA TGIYGGAIYL TGGSMLTSGN
 251 251 LSGVLFVNNS SRSGGAIYAN GNVTFSNNSD LTFQNNNTASP QNSLPAPTTP
 301 301 PTPPAVTPLL GYGGAIFCTP PATPPPPTGVS LTISGENSVT FLENIASEQG
 351 351 GALYGKKISI DSNKSTIIFLQ NTAGKGGAIA IPESGELSLS ANQGDILFNK
 401 401 NLSITSGTPT RNSIHFGKDA KFATLGATQG YTLYFYDPIT SDDLSAASAA
 451 451 ATVVVNPKAS ADGAYSGTIV FSGETLTATE AATPANATST LNQKLELEGG
 501 501 TLALRNGATT NVHNFTQDEK SVVIMDAGTT LATTNGANNT DGAITLNKLV
 551 551 INLDSLGDGTK AAVVNQVQSTN GALTISGTLG LVKNSQDCCD NHGMFNKDLQ
 601 601 QVPILELKAT SNTVTTDFS LGTNGYQQSP YGYQGTWEFT IDTTTHTVTG
 651 651 NWKKTGYLPH PERLAPLIPN SLWANVIDLR AVSQASAADG EDVPGKQLSI
 701 701 TGITNFFFHAN HTGDARSYRH MGGGYLINTY TRITPDAALS LGFGQLFTKS
 751 751 KDYLVLGHGS NVYFATVYNS ITKSLFGSSR FFSGGTSRVFT YSRSENEKVKT
 801 801 SYTKLPKGRC SWSNNCWLGE LEGNLPITLS SRILNLQII PFVKAEVAYA
 851 851 THGGIQENTP EGRIFGHGHL LNVAVPGVVR FGKNSHNRPD FYTIIVAYAP
 901 901 DYYRHNPDCC TTLPINGATW TSIGNNLTRS TLLVQASSHT SVNDVLEIFG
 951 951 HCGCDIRRTS RQYTLDIGSK LRF*

A predicted signal peptide is highlighted.

The cp6736 nucleotide sequence <SEQ ID 4> is:

	1	ATGAAAACGT CTATTCTGAA GTTCTTAATT TCTACCACAC TGGGCCATG
	51	TTTGCTTCAC ACAGCGTTA CTGTAGAAGT TATCATGCCT TCCGAGAACT
5	101	TTGATGGATC GAGTGGGAAG ATTTTCCTT ACACAACACT TTCTGATCCT
	151	AGAGGGACAC TCTGTATTTT TTCAGGGGAT CTCTACATTG CGAATCTTGA
	201	TAATGCCATA TCCAGAACCT CTTCCAGTTG CTTTAGCAAT AGGGCGGGAG
	251	CACTACAAAT CTTAGGAAAA GGTGGGGTTT TCTCTTCTT AAATATCCGT
10	301	TCTTCAGCTG ACGGAGCCGC GATTAGTAGT GTAATCACCC AAAATCCTGA
	351	ACTATGTCCC TTGAGTTTT CAGGATTAG TCAGATGATC TTCGATAACT
	401	GTGAATCTTT GACTTCAGAT ACCTCAGCGA GTAAATGTCAT ACCTCACCGA
	451	TCGGCGATTT ACGCTACAAC GCCCATGCTC TTTCACAAACA ATGACTCCAT
15	501	ACTATTCCAA TACAACCGTT CTGCAGGATT TGGAGCTGCC ATTGAGGCA
	551	CAAGCATCAC AATAGAAAAT ACGAAAAGA GCCTTCTCTT TAATGGTAAT
	601	GGATCCATCT CTAATGGAGG GGCCCTCACG GGATCTGCAG CGATCAACCT
	651	CATCAACAAT AGCCCTCCTG TGATTTCTC AACGAATGCT ACAGGGATCT
	701	ATGGTGGGGC TATTTACCTT ACCGGAGGAT CTATGCTCAC CTCTGGGAAC
	751	CTCTCAGGAG TCTTGTTCGT TAATAATAGC TCGCGCTCAG GAGGGCCTAT
20	801	CTATGCTAAC GGAAATGTCA CATTTCTAA TAACAGCGAC CTGACTTTCC
	851	AAAACAATAC AGCATCTCCA CAAAATCCT TACCTGCACC TACACCTCCA
	901	CCTACACCAC CAGCAGTCAC TCCTTGTAA GGATATGGAG GCGCCATCTT
	951	CTGTACTCCT CCAGCTACCC CCCCACCAAC AGGTGTTAGC CTGACTATAT
	1001	CTGGAGAAAA CAGCGTTACA TTCCTAGAAA ACATTGCCCTC CGAACAAAGGA
	1051	GGAGCCCTCT ATGGAAAAA GATCTCTATA GATTCTAATA AATCTACAAT
25	1101	ATTCTTGGA AATACAGCTG GAAAAGGAGG CGCTATTGCT ATTCCCGAAT
	1151	CTGGGGAGCT CTCTCTATCC GCAAATCAAG GTGATATCCT CTTAACAAAG
	1201	AACCTCAGCA TCACTAGTGG GACACCTACT CGCAATAGTA TTCACTTCGG
	1251	AAAAGATGCC AAGTTGCCA CTCTAGGAGC TACGCAAGGC TATAACCTAT
30	1301	ACTTCTATGA TCCGATTACA TCTGATGATT TATCTGCTGC ATCCGCAGCC
	1351	GCTACTGTGG TCGTCAATCC CAAAGCCAGT GCAGATGGTG CGTATTTCAGG
	1401	GACTATTGTC TTTTCAGGAG AAACCCCTCAC TGCTACCGAA GCAGCAACCC
	1451	CTGCAAATGC TACATCTACA TTAAACCAAA AGCTAGAACT TGAAGGCGGT
	1501	ACTCTCGCTT TAAGAAACGG TGCTACCTTA AATGTTCTATA ACTTCACGCA
	1551	AGATGAAAAG TCCGTCGTCA TCATGGATGC AGGGACCACA TTAGCAACTA
35	1601	CAAATGGAGC TAATAATACT GACGGTGCTA TCACCTTAAA CAAGCTTGTA
	1651	ATCAATCTGG ATTCTTGGA TGGCACTAAA GCGGCTGTGCG TTAATGTCGA
	1701	GAGTACCAAT GGAGCTCTCA CTATATCCGG AACTTTAGGA CTTGTGAAAAA
	1751	ACTCTCAAGA TTGCTGTGAC AACCACGGGA TGTTTAATAA AGATTTCAG
	1801	CAAGTTCCGA TTTAGAACT CAAAGCGACT TCAAATACTG TAACCACTAC
40	1851	GGACTTCAGT CTCGGCACAA ACGGCTATCA GCAATCTCCC TATGGGTATC
	1901	AAGGAACCTG GGAGTTTACCA ATAGACACGA CAACCCATAC GGTACAGGA
	1951	AATTGGAAAA AAACCGGTTA TCTTCCTCAT CCGGAGCGTC TTGCTCCCCCT
	2001	CATTCTTAAT AGCCTATGGG CAAACGTCTA AGATTTACGA GCTGTAAGTC
	2051	AAGCGTCAGC AGCTGATGGC GAAGATGTCC CTGGGAAGCA ACTGAGCATC
45	2101	ACAGGAATTA CAAATTCTT CCATGCGAAT CATAACCGGTG ATGCACCGCAG
	2151	CTACCGCCAT ATGGGTGGAG GCTACCTCAT CAATACCTAC ACACGCATCA
	2201	CTCCAGATGC TCGTTAAGT CTAGGTTTTG GACAGCTGTT TACAAATCT
	2251	AAGGATTACC TCGTAGGTCA CGGTCAATTCT AACGTTTATT TCGCTACAGT
	2301	ATACTCTAAC ATCACCAAGT CTCTGTTGG ATCATCGAGA TTCTCTCAG
	2351	GAGGCACTTC TCGAGTTACC TATAGCCGTA GCAATGAGAA AGTAAAGACT
50	2401	TCATATACAA AATTGCTAA AGGGCGCTGC TCTTGGAGTA ACAATTGCTG
	2451	GTAGGAGAA CTCGAAGGGAA ACCTCCCCAT CACTCTCTCT TCTCGCATCT
	2501	TAAACCTCAA GCAGATCATT CCCTTGTAA AAGCTGAAGT TGCTTACGCG
	2551	ACTCATGGGG GCATCCAAGA AAATACCCCC GAGGGGAGGA TTTTGGACA
55	2601	CGGTCATCTA CTCAACGTT CAGTTCCCGT AGGGCTCCGC TTTGGTAAAAA
	2651	ATTCTCATAA TCGACCAAGAT TTTTACACTA TAATCGTAGC CTATGCTCCT
	2701	GATGTCTATC GTCACAATCC TGATTGCGAT ACGACATTAC CTATTAATGG
	2751	AGCTACGTGG ACCTCTATAG GGAATAATCT AACCGAGAACT ACTTTGCTAG
	2801	TACAAGCATC CAGCCATACT TCAGTAAATG ATGTTCTAGA GATCTCGGG
60	2851	CACTGTGGAT GTGATATCG CAGAACCTCC CGTCAATATA CTCTAGATAT
	2901	AGGAAGCAAA TTACGATTTT AA

The PSORT algorithm predicts an outer membrane location (0.917).

The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 2A, and also as a GST-fusion. Both proteins were used to immunise mice, whose sera were used in a Western blot (Figure 2B) and for FACS analysis (Figure 2C).

The cp6736 protein was also identified in the 2D-PAGE experiment (Cpn0453) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6736 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 3

The following *C.pneumoniae* protein (PID 4376751) was expressed <SEQ ID 5; cp6751>:

10	1	MRFFCFGMLL PFTFVLANEGLQLPLETYITLSPEYQAAPOVGFTHNQNQD
	51	LAIVGNHNDFI LDYKYYRSN GGALTCKNLL ISENIGNVFF EKNVCPNSGG
	101	AIYAAQNCI SKNQNYAFTT NLVSDNPTAT AGSLLGGALF AINCSTNNL
	151	GQGTFVDNLALNKGGALYTE TNLSIKDNKG PIIIKQNRL NSDSLGGGIY
	201	SQNSLNIEGN SGAIQITSSNS SGSGGGIFST QTLTISSNKK LIEISENSAF
15	251	ANNYGSNFNP GGGGLTTFC TILMNREGVL FNNNQSOSNG GAIHAKSIII
	301	KENGPVYFLN NTATRGALL NLSAGSGNGS FILSADNGDI IFNNNTASKH
	351	ALNPPYRNAI HSTPNMNLQI GARPGYRVLF YDPIEHELPSSFPILFNFET
	401	GHTGTIVLFSG EHVNQNFTDE MNFFSYLRNT SELRQGVLAEDGAGLACYK
	451	FFQRGGTLLL GQGAVITTAG TIPTPSSTPT TVGSTITLNH IAIDLPSILS
20	501	FQAQAPKIWI YPTKTGSTYT EDSNPTITIS GTLTLRNSNN EDPYDSDLDS
	551	HSLEKVPLLY IVDVAAQKIN SSQLDLSTLN SGEHYGYQGI WSTYWVETTT
	601	ITNPTSLLGA NTKHKLLIYAN WSPLGYRPHP ERRGEFITNA LWQSAYTALA
	651	GLHSLSSWDE EKGHAASLQIG IGLLIVHQDKD NGFKGFRSHM TGYSATTEAT
	701	SSQSPNFSLG FAQFFSKAKE HESQNSTSSH HYFSGMCIEN TLFKEWIRLS
25	751	VSLAYMFTSE HTHTMYQGLL EGNSQGSFHN HTLAGALSCV FLPQPHGESL
	801	QIYPFITALA IRGNLAAFQE SGDharefsl HRPLTDVSLP VGIRASWKNH
	851	HRVPLVWLTE ISYRSTLYRQ DPELHSKLLI SQGTWTTQAT PVTYNALGIK
	901	VKNTMQVFPK VTLSLDYSA ISSSTLSHYL NVASRMRF*

A predicted signal peptide is highlighted.

30 The cp6751 nucleotide sequence <SEQ ID'6> is:

30	1	ATGCGCTTTT TTTGCTTCGG AATGTTGCTT CCTTTTACTT TTGTATTGGC
	51	TAATGAAGGT CTCCAACCTTC CTTTGGAGAC CTATATTACA TTAAGTCCTG
	101	AATATCAAGC AGCCCCTCAA GTAGGGTTTA CTCATAACCA AAATCAAGAT
	151	CTCGCAATTG TCGGGAAATCA CAATGATTTTC ATCTTGACT ATAAGTACTA
	201	TCGGTCGAAT GGAGGTGCTC TTACCTGTAA GAATCTTCG ATCTCTGAAA
	251	ATATAGGAA TGTCTCTTT GAGAAGAATG TCTGTCCCAA TTCTGGCGGG
	301	GCAATTATG CTGCTAAAAA TTGCACGATC TCCAAGAAC AGAACTATGC
	351	ATTACTACA AACTGGTCT CTGACAATCC TACAGCCACT GCGGGATCAC
	401	TATTGGGTGG AGCTCTCTTT GCCATAAATT GCTCTATTAC TAATAACCTA
	451	GGACAGGGAA CTTTCGTGA CAATCTCGCT TAAATAAGG GGGGTGCCCT
	501	CTATACTGAG ACGAACTTAT CTATTAAGA CAATAAAAGGC CCGATCATAA
	551	TCAAGCAGAA TCGGGCACTA AATTGGACA GTT TAGGAGG AGGGATTAT
	601	AGTGGGAACT CTCTAAATAT AGAGGGAAAT TCTGGAGCTA TACAGATCAC
45	651	AAGCAACTCT TCAGGATCTG GGGGAGGCAT ATTTCTACC CAAACACTCA
	701	CGATCTCTC GAATAAAAAA CTCATAGAAA TCAGTAAAAA TTCCCGCTTC
	751	GCAAATAACT ATGGATCGAA CTTCAATCCA GGAGGAGGAG GTCTTACTAC
	801	CACCTTTGC ACGATATTGA ACAACCGAGA AGGGGTACTC TTTAACAAATA
	851	ACCAAAGCCA GAGCAACGGT GGAGCCATTC ATGCAAATC TATCATTATC
	901	AAAGAAAATG GTCCGTATA CTTTTTAAAT AACACTGCAA CTCGGGGAGG
	951	GGCTCTCCTC AACTTATCAG CAGGTTCTGG AAACGGAAGC TTCATCTTAT
50	1001	CTGCAGATAA TGGAGATATT ATCTTTAAC AATAATACGGC CTCCAAGCAT
	1051	GCCCTCAATC CTCCATACAG AAACGCCATT CACTCGACTC CTAATATGAA
	1101	TCTGCAAATA GGAGCCCGTC CCGGCTATCG AGTGCTGTT TATGATCCCA
	1151	TAGAACATGA GCTCCCTTCC TCCCTCCCCA TACTCTTAA TTTCGAAACC
55	1201	GGTCATACAG GTACAGTTTT ATTTCAGGG GAACATGTAC ACCAGAACTT

1251 TACCGATGAA ATGAATTCTT TTTCTTATTT AAGGAACACT TCGGAACACT
 1301 GTCAGGAGT CTTGCTGTT GAAGATGGTG CGGGGCTGGC CTGCTATAAG
 1351 TTCTTCCAAC GAGGAGGCAC TCTACTTCTA GGTCAAGGTG CGGTGATCAC
 1401 GACAGCAGGA ACGATTCCA CACCATCTC AACACCAACG ACAGTAGGAA
 1451 GTACTATAAC TTTAAATCAC ATTGCCATTG ACCTTCCTTC TATTCTTCT
 1501 TTTCAAGCTC AGGCTCCAAA AATTGGATT TACCCACAA AAACAGGATC
 1551 TACCTATACT GAAGATTCCA ACCCGACAAT CACAATCTCA GGAACCTCTCA
 1601 CCTTACGCAA CAGCAACAAAC GAAGATCCCT ACGATAGTCT GGATCTCTCG
 1651 CACTCTCTTG AGAAAGTCC CCTCTTTAT ATTGTCGATG TCGCTGCACA
 1701 AAAAATTAAC TCTTCGCAAC TGGATCTATC CACATTAAT TCTGGCGAAC
 1751 ACTATGGGTA TCAAGGCATC TGGTCGACCT ATTGGGTAGA AACTACAACA
 1801 ATCACGAACC CTACATCTCT ACTAGGCGCG AATAACAAAC ACAAGCTGCT
 1851 CTATGCAAAC TGGTCTCTC TAGGCTACCG TCCTCATCCC GAACGTCGAG
 1901 GAGAATTCAAT TACGAATGCC TTGTCGCAAT CGGCATATAAC GGCTCTTGCA
 1951 GGACTCCACT CCCTCTCTC CTGGGATGAA GAGAAGGGTC ATGCAGCTTC
 2001 CCTACAAGGC ATTGGTCTTC TGTTTCATCA AAAAGACAAA AACGGTTTA
 2051 AGGGATTCG TAGTCATATG ACAGGTTATA GTGCTACAC CGAACGCAACC
 2101 TCTTCTCAAA GTCCGAATT CTCTTAGGA TTTGTCAGT TCTTCTCCAA
 2151 AGCTAAAGAA CATGAATCTC AAAATAGCAC GTCCTCTCAC CACTATTCT
 2201 CTGGAATGTG CATAGAAAAT ACTCTCTCA AAGAGTGGAT ACGTCTATCT
 2251 GTGTCTCTTG CTTATATGTT TACCTCGGAA CATAACCCATA CAATGTATCA
 2301 GGGTCTCTG GAAGGGAACT CTCAGGGATC TTTCCACAAAC CATAACCTTAG
 2351 CAGGGGCTCT CTCCTGTGTT TTCTTACCTC AACCTCACGG CGAGTCCCTG
 2401 CAGATCTATC CCTTTATAC TGCCTTAGCC ATCCGAGGAA ATCTTGCTGC
 2451 GTTTCAAGAA TCTGGAGACC ATGCTCGGAA ATTTTCCCTA CACCGCCCCC
 2501 TAACGGACGT CTCCCTCCCT GTAGGAATCC GCGCTCTTG GAAGAACAC
 2551 CACCGAGTTTC CCCTAGTCTG GCTCACAGAA ATTTCTATC GCTCTACTCT
 2601 CTATAGGCAA GATCCTGAAC TCCACAGAA ATTACTGATT AGCCAAGGTA
 2651 CGTGGACGAC GCAGGCCACT CCTGTGACCT ACAAATGCTTT AGGGATCAAA
 2701 GTGAAAAATAA CCATGCAGGT GTTCTCTAAA GTCACTCTCT CCTTAGATTA
 2751 CTCTGCGGAT ATTTCTCTC CCACGCTGAG TCACTACTTA AACGTGGCGA
 2801 GTAGAATGAG ATTTTAA

The PSORT algorithm predicts an outer membrane location (0.923).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 3A, 35 and also in his-tagged form. The GST-fusion recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 3B) and for FACS analysis (Figure 3C).

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6751 is a surface-exposed and immunoaccessible protein, and that it 40 is a useful immunogen. These properties are not evident from the sequence alone.

Example 4

The following *C.pneumoniae* protein (PID 4376752) was expressed <SEQ ID 7; cp6752>:

1 MFGMTPAVYS LQTDLSLEKFA LERDEEFRTS FPLLLDSLSTL TGFSPIITTFV
 51 GNRHNSSQDI VLSNYKSIDN ILLLWTSAGG AVSCNNFLLS NVEDHAFFSK
 101 NLAIGTGGAI ACQGACTITK NRGLIIFSN RGLNNASTGG ETRGGAIACN
 151 GDFTISQNQG TFYFVNNSVN NWGGALSTNG HCRIQSNRAP LLFFNNNTAPS
 201 GGGALRSENT TISDNTRPIY FKNNCGNNNG AIQTSVTVAI KNNSGSVIFN
 251 NNTALSGSIN SGNGSGGAIY TTNLSIDDNP GTILFNNNNYC IRDGGAICTQ
 301 FLTIKNSGHV YFTNNQGNWG GALMLLQDST CLLFAEQQNI AFQNNEVFLT
 351 TFGRYNAIHC TPNSNLQLGA NKGYTTAFFD PIEHQHPTTN PLIFNPANAH
 401 QGTILFSSAY IPEASDYENN FISSSKNTSE LRNGVLSIED RAGWQFYKFT
 451 QKGGILKLGH AASIAATTANS ETPSTSVGSQ VIIINNLAINL PSILAKGKAP
 501 TLWIRPLQSS APFTEDNNPT ITLSGPLTLL NEENRDPYDS IDLSEPLQNI
 551 HLLSLSDVTA RHINTDNFHP ESLNATEHYG YQGIWSPYWW ETITTTNNAS
 601 IETANTLYRA LYANWTPFLGY KVNPEYQGDL ATTPLWQSFH TMFSLLRSYN
 651 RTGDSDIERP FLEIQGIADG LFVHQNSIPG APGFRIQSTG YSLQASSETS

-45-

701 LHQKISLGFA QFFTRTKEIG SSNNVSAHNT VSSLYVELPW FQEAFATSTV
 751 LAYGYGDHHL HSLHPSHQEQQ AEGTCYSHTL AAAIGCSFPW QOKSYLHLSP
 801 FVQAIAIRSH QTAFEEIGDN PRKFVSQKPF YNLTLPLGIQ GKWLQSKFHVP
 851 TEWTLELSYQ PVLYQQNPQI GVTLLASGGS WDILGHNYVR NALGYKVHNQ
 901 TALFRSLLDLF LDYQGSVSSS TSTHHLQAGS TLKF*

5 The cp6752 nucleotide sequence <SEQ ID 8> is:

	1	ATGTTCGGGGA	TGACTCCTGC	AGTGTATACT	TTACAAACGG	ACTCCCTTGA
	51	AAAGTTTGCT	TTAGAGAGGG	ATGAAGAGTT	TCGTACGAGC	TTTCCTCTCT
10	101	TAGACTCTCT	CTCCACTCTT	ACAGGATTTT	CTCCAATAAC	TACGTTTGT
	151	GGAATAGAC	ATAAATCCTC	TCAAGACATT	GTACTTTCTA	ACTACAAGTC
	201	TATTGATAAAC	ATCCCTCTTC	TTGGACATC	GGCTGGGGGA	GCTGTGTCT
	251	GTAATAATTT	CTTATTATCA	AATGTTGAAG	ACCATGCCTT	CTTCAGTAAA
	301	AATCTCGCGA	TTGGGACTGG	AGGCCGCGATT	GCTTGCAGG	GAGCCGTGAC
	351	AATCACGAA	AATAGAGGAC	CCCTTATTTT	TTTCAGCAA	CGAGGTCTTA
15	401	ACAATGCGAG	TACAGGAGGA	GAAACTCGTG	GGGGTGGGAT	TGCCTGTAAT
	451	GGAGACTTC	CGATTTCTCA	AAATCAAGGG	ACTTTCTACT	TTGTCACACAA
	501	TTCCGTCAAC	AACTGGGGAG	GAGCCCTCTC	CACCAATGGA	CACTGCCGCA
	551	TCCAAAGCAA	CAGGGCACCT	CTACTCTTTT	TTAACATAC	AGCCCCTAGT
20	601	GGAGGGGGTG	CGCTTCGTAG	TGAAAATACA	ACGATCTCTG	ATAACACGCG
	651	TCCTATTTAT	TTTAAGAAC	ACTGTGGGAA	CAATGGGGG	GCCATTCAAA
	701	CAAGCGTTAC	TGTTGCGATA	AAAAATAACT	CCGGGTGGT	GATTTTCAT
	751	AACAACACAG	CGTTATCTGG	TTCGATAAT	TCAGGAAATG	GTTCAAGGAG
	801	GGCGATTTAT	ACAACAAACC	TATCCATAGA	CGATAACCCCT	GGAACATATT
	851	TTTTCAATAA	TAACTACTGC	ATTGCGGATG	GCGGAGCTAT	CTGTACACAA
25	901	TTTTTGACAA	TCAAAATAG	TGGCCACGTA	TATTTTCACCA	ACAATCAAGG
	951	AAACTGGGGG	GGTGCTCTTA	TGCTCCTACA	GGACAGCACC	TGCCTACTCT
	1001	TCGCGGAACA	AGGAAATATC	GCATTTCAA	ATAATGAGGT	TTTCCTCACC
	1051	ACATTTGGTA	GATACAACGC	CATACATTGT	ACACCAAAATA	GCAACATTACA
	1101	ACTTGGAGCT	AATAAGGGT	ATACGACTGC	TTTTTTGAT	CCTATAGAAC
30	1151	ACCAACATCC	AACTACAAAT	CCTCTAATCT	TTAATCCAA	TGCGAACCAT
	1201	CAGGGAACGA	TCTTATTTTC	TTCAGCCTAT	ATCCCAGAAG	CTTCTGACTA
	1251	CGAAAATAAT	TTCATTAGCA	GCTGAAAAAA	TACCTCTGAA	CTTCGCAATG
	1301	GTGTCTCTC	TATCGAGGAT	CGTGCAGGAT	GGCAATTCTA	TAAGTTCACT
	1351	CAAAAGGAG	GTATCCTAA	ATTAGGGCAT	GCGGCGAGTA	TTGCAACAAAC
35	1401	TGCCAACTCT	GAGACTCCAT	CAACTAGTGT	AGGCTCCAG	GTCATCATT
	1451	ATAACCTTG	GATTAACCTC	CCCTCGATCT	TAGCAAAAGG	AAAAGCTCCT
	1501	ACCTTGTGGA	TCCGTCTCT	ACAATCTAGT	GCTCCTTTCA	CAGAGGACAA
	1551	TAACCCCTACA	ATTACTTTAT	CAGGTCTCT	GACACTCTTA	AATGAGGAAA
	1601	ACCGCGATCC	CTACGACAGT	ATAGATCTCT	CTGAGCCTTT	ACAAAACATT
40	1651	CATCTCTTPT	CTTATCGGA	TGTAACAGCA	CGTCATATCA	ATACCGATAA
	1701	CTTTCATCCT	GAAAGCTTAA	ATGCGACTGA	GCATTACGGT	TATCAAGGCA
	1751	TCTGGTCTCC	TTATTGGGT	GAGACGATAA	CAACAAACAA	TAACGCTTCT
	1801	ATAGAGACGG	CAAACACCCCT	CTACAGAGCT	CTGTATGCCA	ATTGGACTCC
	1851	CTTAGGATAT	AAGGTCAATC	CTGAATACCA	AGGAGATCTT	GCTACGACTC
45	1901	CCCTATGGCA	ATCCTTCAT	ACTATGTTCT	CTCTTAAAG	AAGTTATAAT
	1951	CGAACTGGTG	ATTCTGATAT	CGAGAGGCT	TTCTTAGAAA	TTCAAGGGAT
	2001	TGCCGACGGC	CTCTTTGTT	ATCAAAATAG	CATCCCCGGG	GCTCCAGGAT
	2051	TCCGTATCCA	ATCTACAGGG	TATTCTTAC	AAGCATCTC	CGAAACTTCT
	2101	TTACATCAGA	AAATCTCTT	AGGTTTGCA	CAGTTCTCA	CCCGCACTAA
50	2151	AGAAATCGGA	TCAAGCAACA	ACGTCTCGGC	TCACAATACA	GTCTCTTCAC
	2201	TTTATGTTGA	GCTTCCGTGG	TTCCAAGAGG	CCTTTGCAAC	ATCCACAGTG
	2251	TTAGCGTATG	GCTATGGGG	CCATCACCTC	CACAGCCTAC	ATCCCTCACA
	2301	TCAAGAACAG	GCAGAAGGGG	CGTGTATAG	CCATACATTA	GCAGCAGCTA
	2351	TCGGCTGTT	TTTCCCTTGG	CAACAGAAAT	CCTATCTCA	CCTCAGCCCG
55	2401	TTCGTTCAGG	CAATTGCAAT	ACGTTCTCAC	CAAACAGCGT	TCGAAGAGAT
	2451	TGGTGAACAAT	CCCCGAAAGT	TTGTCTCTCA	AAAGCCTTTC	TATAATCTGA
	2501	CCTTACCTCT	AGGAATCCAA	GGAAAATGGC	AGTCAAAATT	CCACGTACCT
	2551	ACAGAACATGGA	CTCTAGAACT	TTCTTACCAA	CCGGTACTCT	ATCAACAAAA
	2601	TCCCCAAATC	GGTGTACGC	TACTTGCAG	CGGAGGTTCC	TGGGATATCC
60	2651	TAGGCCATAA	CTATGTTCGC	AATGCTTTAG	GGTACAAGT	CCACAATCAA
	2701	ACTGCGCTCT	TCCGTCTCT	CGATCTATT	TTGGATTACC	AAGGATCGGT
	2751	CTCCTCCTCG	ACATCTACGC	ACCATCTCCA	AGCAGGAAGT	ACCTTAAAT
	2801	TCTAA				

The PSORT algorithm predicts a cytoplasmic location (0.138).

The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 4A, and also as a GST-fusion. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (4B) and the his-tagged protein was used for FACS analysis (4C).

The cp6752 protein was also identified in the 2D-PAGE experiment (Cpn0467).

- 5 These experiments show that cp6752 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 5

The following *C.pneumoniae* protein (PID 4376850) was expressed <SEQ ID 9; cp6850>:

10 1 MKRAVLIAAM FCGVVSLSSC CRIVDCCFED PCAPSSCNPC EVIRKKERSC
 51 GGNACGSYVP SCSNPGCGSTE CNSQSPQVKG CTSPDGRCRQ *

A predicted signal peptide is highlighted.

The cp6850 nucleotide sequence <SEQ ID 10> is:

15 1 ATGAAGAAAG CTGTTTTAAC TGCTGCAATG TTTTGTGGAG TAGTTAGCTT
 51 AAGTAGCTGC TGCCGCATTG TAGATTGTG TTTTGAGGAT CCTTGCGCAC
 101 CCTCTTCATTG CAATCCTGT GAAGTAATAA GAAAAAAAAGA AAGATCTTGC
 151 GCGGTTAATG CTTGTGGTC CTACGTTCCCT TCTTGTTCATA ATCCATGTGG
 201 TTCAACAGAG TGTAACTCTC AAAGCCCACA AGTTAAAGGT TGTACATCAC
 251 CTGATGGCAG ATGCAAACAG TAA

The PSORT algorithm predicts an inner membrane location (0.329).

- 20 The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 5A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 5B) and for FACS analysis (Figure 5B). A his-tagged protein was also expressed.

These experiments show that cp6850 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

25 **Example 6**

The following *C.pneumoniae* protein (PID 4376900) was expressed <SEQ ID 11; cp6900>:

30 1 MKIKFSWKVN FLICLLAVGL IFFGCSRVRK EVLVGRDATW FPKQFGIYTS
 51 DTNAFLNDLV SEINYKENLN INIVNQDWHL LFENLDDKKT QGAFTSVLPT
 101 LEMLEHYQFS DPILLTGTVL VVAQDSPYQS IEDLKGRRLIG VYKFDSSVLV
 151 AQNIPDAVIS LYQHVPIALE ALTSNCYDAL LAPVIEVTAL IETAYKGRLK
 201 IISKPLNADG LRLAILKGTN GDLLEGFNAG LVKTRRSRGKY DAIKQRYRLP

The cp6900 nucleotide sequence <SEQ ID 12> is:

35 1 GTGAAGATAA AATTTCTTG GAAAGTAAAT TTTTTAATAT GTTTACTGGC
 51 TGTGGGACTG ATCTTTTCG GGTGCTCTCG AGTAAAAAGA GAAGTTCTCG
 101 TAGGTCGTGA TGCCACCTGG TTTCCAAAAC AATTCGGCAT TTATACATCC
 151 GATACCAACG CATTTTAAAG CGATCTTGTG TCTGAGATTA ACTATAAAGA
 201 GAATCTAAAT ATTAATATTG TAAATCAAGA TTGGGTGCAT CTCTTTGAGA
 251 ATTTAGATGA TAAAAAGACC CAAGGAGCAT TTACATCTGT ATTGCTCTACT
 301 CTTGAGATGC TCGAACACTA TCAATTTCCT GATCCCATT TACTCACAGG
 351 TCTCTGTCCCTT GTCGTCGCTC AAGACTCTCC TTACCAATCT ATAGAGGATC
 401 TTAAAGGTGCG TCTTATTGGT GTGTATAAGT TTGACTCTTC AGTTCTTGTG
 451 GCTCAAAATA TCCCTGACGC TGTGATTAGC CTCTACCAAC ATGTTCCAAT
 501 AGCATTGGAA GCCTTAACAT CGAATTGTTA CGACGCTCTT CTAGCTCCTG
 551 TAATTGAAGT GACCGCGCTA ATAGAAACAG CATATAAAGG AAGACTGAAA
 601 ATTATTTCAA AACCTTAAAG CGCAGATGGT TTGGGGCTTG CAATACTGAA

-47-

651 AGGGACAAAC GGAGATTTGC TTGAAGGGTT TAACGCAGGA CTTGTGAAAA
 701 CACGACGCTC AGGAAAATAC GATGCTATAA AACAGCGGTA TCGTCTTCCC
 751 TAA

The PSORT algorithm predicts an inner membrane location (0.452).

- 5 The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 6A. The recombinant protein was used to immunise mice, whose sera were used for FACS analysis (Figure 6B). A his-tagged protein was also expressed.

The cp6900 protein was also identified in the 2D-PAGE experiment (Cpn0604).

- 10 These experiments show that cp6900 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 7

The following *C.pneumoniae* protein (PID 4377033) was expressed <SEQ ID 13; cp7033>:

1 MVNPIGPGPI DETERTPPAD LSAQGLEASA ANKSAEAQRI AGAEAKPKES
 51 KTDSSVERWSI LRSAVNALMS LADKLGIASS NSSSSSTSRSVA DVDSSTTATAP
 101 TPPPPPTFDDY KTQQAQTAYDT IFTSTSLADI QAALVSLQDA VTNIKDSTAAT
 151 DEETAIAAAEW ETKNADAVKV GAQITELAKY ASDNQAIILDS LGKLTSDLL
 201 QAALLQSVAN NNKAAELLKE MQDNPVVPVGK TPAIAQSLVD QTDATATQIE
 251 KDGNAIRDAY FAGQNASGAV ENAKSNNNS NIDSAAIA TAKTQIAEAQ
 301 KKFPDPSPILQ EAEQMVIIQAE KDLKNIKPAD GSDVPNPGTT VGGSKQQGSS
 351 IGSIRVSMILL DDAENETASI LMSGFRQMIH MFNTENPDSQ AAQQELAAQA
 401 RAAKAAGDDS AAAALADAQK ALEAALGKAG QQQGILNALG QIASAAVVSA
 451 GVPAAASSI GSSVKQLYKT SKSTGSDYKT QISAGYDAYK SINDAYGRAR
 501 NDATRDVINN VSTPALTRSV PRARTEARGP EKTDQALARV ISGNNSRTLGD
 551 VYSQVSALQS VMQIIQSNPQ ANNEEIROQL TSAVTKPPQF GYPYVQLSND
 601 STQKFIAKLE SLFAEGSRTA AEIKALSFET NSLFIQQVLV NIGSLYSGYL
 651 Q*

The cp7033 nucleotide sequence <SEQ ID 14> is:

1 ATGGTTAACATC CTATTGGTCC AGGTCCCTATA GACGAAACAG AACGCCACACC
 51 TCCCGCAGAT CTTTCTGCTC AAGGATTGGA GCGGAGTGCA GCAAATAAGA
 101 GTGCGGAAGC TCAAAGATA GCAGGTGCGG AAGCTAAGCC TAAAGAATCT
 151 AAGACCGATT CTGTAGAGCG ATGGAGCATC TTGCGTTCTG CAGTGAATGC
 201 TCTCATGAGT CTGGCAGATA AGCTGGGTAT TGCTTCTAGT AACAGCTCGT
 251 CTTCTACTAG CAGATCTGCA GACGTGGACT CAACGACAGC GACCCACCT
 301 ACGCCCTCC CACCCACCGTT TGATGATTAT AAGACTCAAG CGCAAACAGC
 351 TTACGATACT ATCTTTACCT CAACATCACT AGCTGACATA CAGGCTGCTT
 401 TGGTGAGCCT CCAGGATGCT GTCACTAATA TAAAGGATAC AGCGGCTACT
 451 GATGAGGAAA CCGCAATCGC TGCGGAGTGG GAAACTAAGA ATGCCGATGC
 501 AGTTAAAGTT GGCAGCAGAA TTACAGAATT AGCGAAATAT GCTTCGGATA
 551 ACCAAGCGAT TCTTGACTCT TTAGGTAAAC TGACTTCCCT CGACCTCTTA
 601 CAGGCTGCTC TTCTCAACATC TGTAGCAAAC AATAACAAAG CAGCTGAGCT
 651 TCTTAAAGAG ATGCAAGATA ACCCAGTAGT CCCAGGGAAA ACGCCTGCAA
 701 TTGCTCAATC TTTAGTTGAT CAGACAGATG CTACAGCCAC ACAGATAGAG
 751 AAAGATGGAA ATGCGATTAG GGATGCATAT TTTGCAGGAC AGAACGCTAG
 801 TGGAGCTGTA GAAAATGCTA AATCTAATAA CAGTATAAGC AACATAGATT
 851 CAGCTAAAGC AGCAATCGCT ACTGCTAAGA CACAAATAGC TGAAGCTCAG
 901 AAAAAGTTCC CCGACTCTCC AATTCTCAA GAAGCGGAAAC AAATGGTAAT
 951 ACAGGCTGAG AAAGATCTTA AAAATATCAA ACCTGCAGAT GGTTCTGATG
 1001 TTCCAATCC AGGAACCTACA GTTGGAGGCT CCAAGCAACA AGGAAGTAGT
 1051 ATTGGTAGTA TTCGTTGTTTC CATGCTGTTA GATGATGCTG AAAATGAGAC
 1101 CGCTTCCATT TTGATGTCG GGTTTCTGCA GATGATTCAAC ATGTTCAATA
 1151 CGGAAAATCC TGATTCTCAA GCTGCCAAC AGGAGCTCGC AGCACAAGCT
 1201 AGAGCAGCGA AAGCCGCTGG AGATGACAGT GCTGCTGCAG CGCTGGCAGA
 1251 TGCTCAGAAA GCTTTAGAAG CGGCTCTAGG TAAAGCTGGG CAACAAACAGG
 1301 GCATACTCAA TGCTTTAGGA CAGATCGCTT CTGCTGCTGT TGTGAGCGCA
 1351 GGAGTTCCCTC CCGCTGCAGC AAGTTCTATA GGGTCATCTG TAAAACAGCT
 1401 TTACAAGAGACC TCAAAATCTA CAGGTTCTGA TTATAAAACA CAGATATCAG

5 1451 CAGGTTATGA TGCTTACAAA TCCATCAATG ATGCCTATGG TAGGGCACGA
 1501 AATGATGCGA CTCGTGATGT GATAAACAAAT GTAAAGTACCC CCGCTCTCAC
 1551 ACGATCCGTT CCTAGAGCAC GAACAGAAC TCGAGGACCA GAAAAAACAG
 1601 ATCAAGCCCT CGCTAGGGTG ATTTCTGGCA ATAGCAGAAC TCTTGGAGAT
 1651 GTCTATAGTC AAGTTTCGGC ACTACAATCT GTAAATGCAGA TCATCCAGTC
 1701 GAATCCTCAA GCGAATAATG AGGAGATCG ACAAAAGCTT ACATCGGCAG
 1751 TGACAAAGCC TCCACAGTT GGCTATCCTT ATGTGCAACT TTCTAATGAC
 1801 TCTACACAGA AGTCATAGC TAAATTAGAA AGTTTGTGCT CTGAAGGATC
 1851 TAGGACAGCA GCTGAAATAA AAGCACTTTC CTTTGAAACG AACTCCTTGT
 1901 TTATTCAGCA GGTGCTGGTC AATATCGGCT CTCTATATTC TGGTTATCTC
 1951 CAATAA

The PSORT algorithm predicts a cytoplasmic location (0.272).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 7A. A his-tagged protein was also expressed. The recombinant proteins were used to immunise mice, whose
 15 sera were used for FACS (Figure 7B) and Western blot (7C) analyses.

The cp7033 protein was also identified in the 2D-PAGE experiment (Cpn0728) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7033 a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

20 Example 8

The following *C.pneumoniae* protein (PID 6172321) was expressed <SEQ ID 15; cp0017>:

25 1 MGIKGTGIIIV WVDDATAKTK NATLTWTKTG YKPNPERQGP LVPNSLWGSF
 51 VDVRSIQSLM DRSTSSLSSS TNLWVSGIAD FLHEDQKGNO RSYRHSSAGY
 101 ALGGGFTTAS ENFFNFACFCQ LFGYDKDHLY AKNHHTHVYAG AMSYRHLGES
 151 KTLAKILSGN SDSLPFVFNA RFAYGHTDNM MTTKYTGSP VKGSGWNDAP
 201 GIECGGAIPV VASGRRSWVD THTPFLNLEM IYAHQNDFKE NGTEGRSRFQS
 251 EDLFNLAVPV GIKFEKFSDK STYDLSIAYV PDVIRNDPGC TTTLMVSGDS
 301 WSTCGTSLSR QALLVRAGNH HAFASNFEVF SQFEVELRGS SRSYAIDLGG
 351 RFGF*

30 The cp0017 nucleotide sequence <SEQ ID 16> is:

35 1 ATGGGTATCA AGGGAACCTGG AATAATTGTT TGGGTCGACG ATGCAACTGCG
 51 AAAAACAAAAA AATGCTAACCT TAACCTTGGAC TAAAACAGGA TACAAGCCGA
 101 ATCCAGAACG TCAGGGACCT TTGGTTCCCA ATAGCCTGTG GGGTTCTTTT
 151 GTCGATGTCC GCTCCATTCA GAGCCTCATG GACCGGAGCA CAAGTTCGTT
 201 ATCTTCGTCA ACAAAATTGTT GGGTATCAGG AATCGCGGAC TTTTTGCATG
 251 AAGATCAGAA AGGAAACCAA CGTAGTTATC GTCAATTCTAG CGCGGGTTAT
 301 GCATTAGGAG GAGGATTCTT CACGGCTTCT GAAAATTCT TTAATTTCGC
 351 TTTTTGTCACT CTTTTGGCT ACGACAAGGA CCATCTGTG GCTAAGAACCC
 401 ATACCCATGT ATATGCAGGG GCAATGAGTT ACCGACACCT CGGAGAGTCT
 451 AAGACCCCTCG CTAAGATTTT GTCAAGGAAAT TCTGACTCCC TACCTTTGT
 501 CTTCATGCT CGGTTTGCTT ATGCCATAC CGACAATAAC ATGACCACAA
 551 AGTACACTGG CTATTCTCCT GTTAAGGGAA GCTGGGGAAA TGATGCCTTC
 601 GGTATAGAAT GTGGAGGAGC TATCCCCGTA GTTGCTTCAG GACGTCGGTC
 651 TTGGGTGGAT ACCCACACGC CATTCTAAA CCTAGAGATG ATCTATGCAC
 701 ATCAGAATGA CTTTAAGGAA AACGGCACAG AAGGCCGTTT TTTCCAAAGT
 751 GAAGACCTCT TCAATCTAGC GGTCTCTGTA GGGATAAAAT TTGAGAAATT
 801 CTCCGATAAG TCTACGTATG ATCTCTCCAT AGCTTACGTT CCCGATGTGA
 851 TTGCGTAATGA TCCAGGCTGC ACGACAACTC TTATGGTTT TGGGGATTCT
 901 TGGTCGACAT GTGGTACAAG CTTGTCTAGA CAAGCTCTTC TTGTCAGTGC
 951 TGGAAATCAT CATGCCTTG CTTCAAACCT TGAAGTTTC AGTCAGTTG
 1001 AAGTCGAGTT GCGAGGTTCT TCTCGTAGCT ATGCTATCGA TCTTGGAGGA
 1051 AGATTCCGGAT TTTAA

This sequence is frame-shifted with respect to cp0016.

The PSORT algorithm predicts a cytoplasmic location (0.075).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 8A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 8B) and for FACS analysis (Figure 8C). A his-tagged protein was also expressed.

- 5 This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp0017 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 9

- 10 The following *C.pneumoniae* protein (PID 6172315) was expressed <SEQ ID 17; cp0014>:

```

1  MKSSFPKFVF STFAIFFPLSM IATETVLDSS ASFDGNKNGN FSVRESQEDA
51  GTTYLFKGKV TLENIPGTGT AITKSCFNNT KGDLTFTGNG NSLLFQTVDA
101 GTVAGAAVNS SVVDKSTTFI GFSSLSFIAS PGSSITTGKG AVSCSTGSL
151 LTKMSVCSSA KTFQRIMAVL SPQKLFH*

```

- 15 The cp0014 nucleotide sequence <SEQ ID 18> is:

```

1  ATGAAGTCCTT CTTTCCCCAA GTTTGTATTT TCTACATTG CTATTTTCCC
51  TTTGTCTATG ATTGCTACCG AGACAGTTT GGATTCAAGT GCGAGTTTCG
101 ATGGGAATAA AAATGGTAAT TTTTCAGTTC GTGAGAGTCA GGAAGATGCT
151 GGAACACTACCT ACCTATTAA GGGAAATGTC ACTCTAGAAA ATATTCCCTGG
20  201 AACAGGCACA GCAATCACAA AAAGCTGTT TAACAACACT AAGGGCGATT
251 TGACTTTACAC AGGTAACGGG AACTCTCTAT TGTTCCAAAC GGTGGATGCA
301 GGGACTGTAG CAGGGCTGC TGTTAACAGC AGCGTGGTAG ATAAATCTAC
351 CACGTTTATA GGGTTTCTT CGCTATCTTT TATTGCGTCT CCTGGAAGTT
401 CGATAACTAC CGGCAAAGGA GCCGTTAGCT GCTCTACGGG TAGCTTGAGT
451 TTGACAAAAAA TGTCAGTTTG CTCITTCAGCA AAAACTTTTC AACGGATAAT
501 GGCGGTGCTA TCACCGCAAA AACTCTTCA TTAA

```

This protein is frame-shifted with respect to cp0015.

The PSORT algorithm predicts an inner membrane location (0.047).

- 30 The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 9A. A GST-fusion was also expressed. The recombinant proteins were used to immunise mice, whose sera were used in an immunoassay (Figure 9B) and for FACS analysis (Figure 9C).

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments suggest that cp0014 is a useful immunogen. These properties are not evident from the sequence alone.

Example 10

- The following *C.pneumoniae* protein (PID 6172317) was expressed <SEQ ID 19; cp0015>:

```

1  MSALFSENTS SKKGGAIQTS DALTITGNQG EVSFSDNTSS DSGAAIFTEA
51  SVTISNNNAKV SFIDNKVTGA SSSTTGDMMSG GAICAYKTST DTKVTLTGNQ
101 MLLFSNNST TAGGAIYVKK LELASGLTL FSRNSVNGGT APKGGAIIAIE
151 DSGELSLSAD SGDIVFLGNT VTSTTPGTNR SSIDLGTSAK MTALRSAAGR

```

-50-

5 201 AIYFYDPITT GSSTTVTDVL KVNETPADSA LOYTGNIIIFT GEKLSSETAA
 251 DSKNLTSKLL QPVTLGGT SLKHGVTLQT QAFTQQADSR LEMDVGTITLE
 301 PADTSTINNL VINISSIDGA KKAKIETKAT SKNLTLSGTI TLLDPGTIFY
 351 ENHSLRNPQS YDILELKASG TVTSTAVTPD PIMGEKFHYG YQGTWGPIVW
 401 GTGASTTATF NWTKTGIPN PERIGSLVPN SLWNAFIDIS SLHYLMETAN
 451 EGLQGDRAFW CAGLSNFHK DSTKTRRGFR HLSGGYVIGG NLHTCSDKIL
 501 SAAFCQLFGR DRDYFVAKNQ GTVYGGTLYY QHNETYISLP CKLRPCSLSY
 551 VPTEIPVLFS GNLSYTHTDN DLKTKYTTYP TVKGSGWNDS FALEFGGRAP
 601 ICLDESALFE QYMPFMKLQF VYAHQEGFKE QGTEAREFGS SRLVNLLALPI
 651 GIRFDKESDC QDATYNLTG YTVDLVRSNP DCTTTLRISG DSWKTFGTNL
 701 ARQALVLRAG NHFCFNSNFE AFSQSFELR GSSRNVNVDL GAKYQF*

This sequence is frame-shifted with respect to cp0014.

The cp0015 nucleotide sequence <SEQ ID 20> is:

15 1 ATGTCAGCTC TGTTTCTGA AAATACCTCC TCAAAGAAAAG GCGGAGCCAT
 51 TCAGACTTCC GATGCCCTTA CCATTACTGG AAACCAAGGG GAAGTCTCTT
 101 TTTCTGACAA TACTTCTCG GATTCTGGAG CTGCAATTTC TACAGAACCC
 151 TCGGTGACTA TTTCTAAATA TGCTAAAGTT TCCTTTATTG ACAATAAGGT
 201 CACAGGAGCG AGCTCCTCAA CAACGGGGGA TATGTCAGGA GGTGCTATCT
 251 GTGCTTATAA AACTAGTACA GATACTAAGG TCACCCCTCAC TGGAAATCAG
 301 ATGTTACTCT TCAGCAACAA TACATCGACA ACAGCGGGAG GAGCTATCTA
 351 TGTGAAAAG CTCGAACCTGG CTTCCGGAGG ACTTACCCCTA TTCAGTAGAA
 401 ATAGTGTCAA TGGAGGTACA GCTCCTAAAG GTGGAGCCAT AGCTATCGAA
 451 GATAGTGGGG ATTGAGTT ATCCGCCGAT AGTGGTGACA TTGTCTTTTT
 501 AGGGAATACA GTCACTTCTA CTACTCCCTGG GACGAATAGA AGTAGTATCG
 551 ACTTAGGAAC GAGTGCAAAG ATGACAGCTT TGCGTCTGC TGCTGGTAGA
 601 GCCATCTACT TCTATGATCC CATAACTACA GGATCATCCA CAACAGTTAC
 651 AGATGTCTTA AAAGTTAATG AGACTCCGGC AGATTCTGCA CTACAATATA
 701 CAGGGAACAT CATCTTCACA GGAGAAAAGT TATCAGAGAC AGAGGCCGCA
 751 GATTCTAAAA ATCTTACTTC GAAGCTACTA CAGCCTGTAA CTCTTTCTAGG
 801 AGGTAACCTA TCTTTAAAAC ATGGAGTGAC TCTGCAGACT CAGGCATTCA
 851 CTCAACAGGC AGATTCTCGT CTCGAATATGG ACCTGAGAAC TACTCTAGAA
 901 CCTGCTGATA CTAGCACCAT AAACAATTTC GTCTTAACA TCAGTTCTAT
 951 AGACGGTGCA AAGAAGGCAA AAATAGAAAAC CAAAGCTACG TCAAAAAATC
 1001 TGACTTTATC TGGAACCCATC ACTTTATTGG ACCCGACGGG CACGTTTTAT
 1051 GAAAATCATA GTTTAAGAAA TCCTCAGTCC TACGACATCT TAGAGCTCAA
 1101 AGCTTCTGGA ACTGTAACAA GCACCGCAGT GACTCCAGAT CCTATAATGG
 1151 GTGAGAAATT CCATTACGGC TATCAGGGAA CTTGGGGCCC ATTGTTGG
 1201 GGGACAGGGG CTTCTACGGC TGCAACCTTC AACTGGACTA AAACTGGCTA
 1251 TATTCTTAAT CCCGAGCGTA TCGGCCTCTTT AGTCCCTAAT AGCTTATGGA
 40 1301 ATGCATTATAG ATGATATTAGC TCTCTCCATT ATCTTATGGA GACTGCAAAC
 1351 GAAGGGTTGC AGGGAGACCG TGCTTTTTGG TGTGCTGGAT TATCTAACTT
 1401 CTTCCATAAG GATAGTACAA AAACACGACG CGGGTTTCGC CATTGAGTG
 1451 GCGGTTATGT CATAGGAGGA AACCTACATA CTTGTTCA TAAGATTCTT
 1501 AGTGCTGCAT TTGTCAGCT CTTTGGAAAGA GATAGAGACT ACTTTGTAGC
 45 1551 TAAAGATCAA GGTACAGTCT ACGGAGGAAC TCTCTATTAC CAGCACAACG
 1601 AAACCTATAT CTCTCTTCT TGCAAACATAC GGCCTTGTC GTGCTTTAT
 1651 GTTCCTACAG AGATTCTCGT TCTCTTTCA GGAAACCTTA GCTACACCCA
 1701 TACGGATAAC GATCTGAAAA CCAAGTATAC AACATATCT ACTGTTAAAG
 1751 GAAGCTGGGG GAATGATAGT TTCGCTTTAG AATTGGTGG AAGAGCTCCG
 50 1801 ATTTGCTTAG ATGAAAGTGC TCTATTGAG CAGTACATGC CCTTCATGAA
 1851 ATTGCAAGTT GTCTATGCAC ATCAGGAAGG TTTTAAAGAA CAGGGAACAG
 1901 AAGCTCGTGA ATTGGAAGT AGCCGTCTTG TGAATCTTC CTTACCTATC
 1951 GGGATCCGAT TTGATAAGGA ATCAGACTGC CAAGATGCAA CGTACAATCT
 2001 AACTCTTGGT TATACTGTCG ATCTTGTTCG TAGTAACCCC GACTGTACGA
 2051 CAAACACTGCG ATTAGCGGT GATTCTTGGA AAACCTTCGG TACGAATTG
 2101 GCAAGACAAG CTTTAGCTC TCGTGCAGGG AACCATTTT GCTTTAACTC
 2151 AAATTTGAA GCCTTTAGCC AATTCTCTTT TGAATTGCGT GGGTCATCTC
 2201 GCAATTACAA TGTAGACTTA GGAGCAAAAT ACCAATTCTA A

The PSORT algorithm predicts a cytoplasmic location (0.274).

- 60 The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 10A.
 The recombinant protein was used to immunise mice, whose sera were used in a Western blot
 (Figure 10B) and for FACS analysis. A his-tagged protein was also expressed.

These experiments show that cp0015 is a useful immunogen. These properties are not evident from the sequence alone.

Example 11

The following *C.pneumoniae* protein (PID 6172325) was expressed <SEQ ID 21; cp0019>:

```

5      1 LQDSQDYSFV KLSPGAGGTI ITQDASQKPL EVAPSRPHYG YQGHWNVQVI
      51 PGTGTQPSQA NLEWVRTGYL PNPERQGSLV PNSLGWSFVD QRAIQEIMVN
     101 SSQILCQERG VWGAGIANFL HRDKINEHGY RHSGVGVLVG VGTHAFSDAT
     151 INAAFCQLFS RDKDYVVSKN HGTSYSGVVF LEDTLEFRSP QGFYTDSSSE
     201 ACCNQVVTID MQLSYSHRNN DMKTKYTTYP EAQGSWANDV FGLEFGATTY
     251 YYPNSTFLFD YYSPFLLQC TYAHQEDFKE TGGEVRHFTS GDLFNLAVPI
     301 GVKFERFSDC KRGSYELTLA YVPDVIRKDP KSTATLASGA TWSTHGNNL
     351 RQGLQLRLGN HCLINPGIEV FSHGATELRG SSRNNYNINLG GKYRF*

```

This sequence is frame-shifted with respect to cp0018.

The cp0019 nucleotide sequence <SEQ ID 22> is:

```

15     1 TTGCAAGACT CTCAGACTA TAGCTTGTA AAGTTATCTC CAGGAGCGGG
      51 AGGGACTATA ATTACTCAAG ATGCTTCTCA GAAGCCTCTT GAAGTAGCTC
     101 CTTCTAGACC ACATTTATGGC TATCAAGGAC ATTGGAATGT GCAAGTCATC
     151 CCAGGAACGG GAACTCAACC GAGCCAGGCA AATTTAGAAT GGGTGCAGAC
     201 AGGATACCTT CCGAATCCCAG AACGGCAAGG ATCTTATGTT CCCAATAGCC
     251 TGTGGGGTTTC TTTTGTGAT CAGCGTGCTA TCCAAGAAAT CATGGTAAAT
     301 AGTAGCCAAA TCTTATGTCA GGAACGGGGAG GTCTGGGGAG CTGGAATTGC
     351 TAATTTCCTA CATAGAGATA AAATTAATGA GCACGGCTAT CGCCATAGCG
     401 GTGTCGGTTA TCTTGTTGGGA GTTGGCACTC ATGCTTTTTC TGATGCTACG
     451 ATAATGCGG CTTTTGCCA GCTCTTCAGT AGAGATAAAG ACTACGTAGT
     501 ATCCAAAAAT CATGGAACTA GCTACTCAGG GGCGTATTCTT CTTGAGGATA
     551 CCCTAGAGTT TAGAAGTCCA CAGGGATTCT ATACTGATAG CTCCCTCAGAA
     601 GCTTGCTGTA ACCAAGTCGT CACTATAGAT ATGCAGTTGT CTTACAGCCA
     651 TAGAAATAAT GATATGAAAAA CCAAATACAC GACATATCCA GAAGCTCAGG
     701 GATCTTGGGC AAATGATGTT TTTGGTCTTG AGTTGGAGC GACTACATAC
     751 TACTACCCCTA ACAGTACTTT TTTATTGAT TACTACTCTC CGTTTCTCAG
     801 GCTGCAGTGC ACCTATGCTC ACCAGGAAGA CTTCAAAGAG ACAGGAGGTG
     851 AGGTTCGTCA CTTTACTAGC GGAGATCTTT TCAATTTCAG AGTTCCATT
     901 GGCCTGAAGT TTGAGAGATT TTCAGACTGT AAAAGGGGAT CTTATGAACT
     951 TACCCCTTGCT TATGTTCTG ATGTGATTG CAAAGATCCC AAGAGCACGG
    1001 CAACATTGGC TAGTGGAGCT ACCTGGAGCA CCCACGGAAA CAATCTCTCC
    1051 AGACAAGGGAT TACAACGTGC TTTAGGGAAC CACTGTCTCA TAAATCCTGG
    1101 AATTGAGGTG TTCAGTCACG GAGCTATTGA ATTGCGGGGA TCCTCTCGTA
    1151 ATTATAACAT CAATCTCGGG GGTAAATACC GATTTTAA

```

The PSORT algorithm predicts a cytoplasmic location (0.189).

40 The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 11A. This protein was used to immunise mice, whose sera were used in a Western blot (Figure 11B) and an immunoblot assay (Figure 11C). A his-tagged protein was also expressed.

These experiments show that cp0019 is a useful immunogen. These properties are not evident from the sequence alone.

Example 12

The following *C.pneumoniae* protein (PID 4376466) was expressed <SEQ ID 23; cp6466>:

```

50     1 MRKISVGICL TILLSLSVVL QGCKESSHSS TSRGELAINI RDEPRSLDPR
      51 QVRLLSEISL VKHIYEGLVQ ENNLSGNIEP ALAEDYSISSL DGLTYTFKLK
     101 SAFWSNGDPL TAEDFIESWK QVATQEVSIGI YAFALNPIKN VRKIQEGHLS
     151 IDHFGVHSPN ESTLTVVLTES PTSHFLKLLA LPVFFPVHKS QRTLQSKSLP
     201 IASGAFYPKN IKQKQWIKLS KNPHYYNQSQ VETKTITIHF IPDANTAAKL

```

5
251 FNQGKLNWQG PPWGERIPQE TLSNLQSKGH LHSFDVAGTS WLTFNINKFP
301 LNNMKLREAL ASALDKEALV STIFLGRAKT ADHLLPTNIH SYPEHQKQEM
351 AQRQAYAKKL FKEALEELQI TAKDLEHLNL IFFVSSSASS LLVQLIREQW
401 KESLGFAIPI VGKEFALLQA DLSSGNFSLA TGGWFADFAD PMAFLTIFAY
451 PSGVPPYAIN HKDFLEILQN IEQEQDHQKR SELVSQASLY LETFHJIIEPI
501 YHDAFQFAMN KKLSNLGVSP TGVVDFRYAK EN*

A predicted signal peptide is highlighted.

The cp6466 nucleotide sequence <SEQ ID 24> is:

10
1 ATCGCGCAAGA TATCAGTGGG AATCTGTATC ACCATTCTCC TTAGCCTCTC
51 CGTAGTCCTC CAAGGCTGCA AGGAGTCCAG TCACCTCTCT ACATCTCGGG
101 GAGAACTCGC TATTAATATA AGAGATGAAC CCCGTTCTTT AGATCCAAGA
151 CAAGTGCAGC TTCTTCAGA AATCAGCCTT GTCAAACATA TCTATGAGGG
201 ATTAGTTCAA GAAAATAATC TTTCAGGAAA TATAGAGCCT GCTCTTGCAG
251 AAGACTACTC TCTTTCTCG GACGGACTCA CTTATACTTT TAAACTGAAA
15
301 TCAGCTTTTG GGAGTAATGG CGACCCCTTA ACAGCTGAAG ACTTTATAGA
351 ATCTTGGAAA CAAGTAGCTA CTCAAGAAGT CTCAGGAATC TATGCTTTG
401 CCTTGAATCC AATTAAAAAT GTACGAAAGA TCCAAGAGGG ACACCTCTCC
451 ATAGACCATT TTGGAGTGCA CTCTCCTAAT GAATCTACAC TTGTTGTTAC
20
501 CCTTGAATCC CCAACCTCGC ATTCTTAAA ACTTTTAGCT CTTCCAGTCT
551 TTTTCCCCGT TCATAAAATCT CAAAGAACCC TGCAATCCAA ATCTCTACCT
601 ATAGCAAGCG GAGCTTTCTA TCCTAAAAAT ATCAAACAAA ACAATGGAT
651 AAAACTCTCA AAAAACCCCTC ACTACTATAA TCAAAGTCAG GTGAAACTA
701 AAACGATTAC GATTCACTTC ATTCCCGATG CAAACACAGC AGCAAAACTA
751 TTTAATCAGG GAAAACCTCAA TTGGCAAGGA CCTCCTTGGG GAGAACGCAT
25
801 TCCTCAAGAA ACCCTATCCA ATTACAGTC TAAGGGGCAC TTACACTCTT
851 TTGATGTCGC AGGAACCTCA TGGCTCACCT TCAATATCAA TAAATTCCCC
901 CTCAACAAATA TGAAGCTTAG AGAACCTTA GCATCAGCCT TAGATAAGGA
951 AGCTCTTGTGTC TCAACTATAT TCTTAGGCCG TGCAAAACT GCCGATCATC
30
1001 TCTTACCTAC AAATATTCTAT AGCTATCCCG AACATCAAA ACAAGAGATG
1051 GCACAACGCG AAGCTTACGC TAAAAAAACTC TTTAAAGAAG CTTTAAAGAAG
1101 ACTCCAAATC ACTGCTAAAG ATCTCGAACAA TCTTAATCTT ATCTTTCCCG
1151 TTCTCTCGTC AGCAAGTTCT TTACTAGTCC AACTTATACG AGAACAGTGG
1201 AAAGAAAGTT TAGGGTTTCGC TATCCTTATT GTCGGAAAGG AATTGCTCT
1251 TCTCCAAGCA GACCTATCTT CAGGGAACCTT CTCTTTAGCT ACAGGAGGAT
35
1301 GGTTCGCGAGA CTTTGCTGAT CCTATGGCAT TTCTAACGAT CTTTGCTTAT
1351 CCATCAGGAG TTCCCTCTTA TGCAATCAAC CATAAGGACT TCCTAGAAAT
1401 TCTACAAAAC ATAGAACAAAG AGCAAGATCA CCAAAACGC TCAGGATTAG
1451 TGTCGCAAGC TTCTCTTAC CTAGAGACCT TTCATATTAT TGAGCCGATC
1501 TACCAACGACG CATTTCATT TGCTATGAAT AAAAAACTTT CTAATCTAGG
40
1551 AGTCTCACCA ACAGGAGTTG TGGACTTCCG TTATGCTAAG GAAAATTAG

The PSORT algorithm predicts that the protein is an outer membrane lipoprotein (0.790).

The protein was expressed in *E.coli* and purified both as a GST-fusion product and a His-tag fusion product. Purification of the protein as a GST-fusion product is shown in Figure 12A. The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 45 12B and 12C). FACS analysis was also performed.

These experiments show that cp6466 is a useful immunogen. These properties are not evident from the sequence alone.

Example 13

The following *C.pneumoniae* protein (PID 4376468) was expressed <SEQ ID 25; cp6468>:

50
1 MFSRWITLFL LFISLTGCSS YSSKHQKSLI IPIHDDPVAF SPEQAKRAMD
51 LSIAQLLFDG LTRETHRESN DLELAIASRY TVSEDFCSYT FFIKDSALWS
101 DGTPTITSEDI RNAWEYAQEN SPHIQIFQGL NFSTPSSNAI TIHLDSPNPD
151 FPKLLAFTPAP AIFKPENPKL FSGPYTLVEY FPGHNIHLKK NPNEYDYHCV
201 SINSIKLLII PDIYTAIHLI NRGKVDWVGQ PWHQGIPWEL HKQSQYHYYT
55
251 YPVEGAFWLC LNTKSPHLND LQNRHRRLATC IDKRSIIIEEA LQGTQQPAET

-53-

301 LSRGAPQPNQ YKKQKPLTPQ EKLVLVTPSD ILRCQRIAEI LKEQWKAAGI
 351 DLILEGLEYH LFVNKRKVQD YAIATQTGVA YYPGANLISE EDKLLQNFEI
 401 IPIYYLSYDY LTQDFIEGVI YNASGAVDLK YTYFP*

A predicted signal peptide is highlighted.

5 The cp6468 nucleotide sequence <SEQ ID 26> is:

1	ATGTTTCAC GATGGATCAC CCTCTTTTA TTATTCATTA GCCTTACTGG
51	ATGCTCCTCC TACTCTCAA AACATAAACAA ATCTTTAATT ATTCCCATAC
101	ATGACGACCC TGTAGCTTT TCTCCTGAAC AAGCAAAACG GGCCATGGAC
151	CTTTCTATTG CCCAACTCT TTTTGATGGT CTGACTAGAG AAACCTCATCG
201	CGAATCCAAT GATTGGAAT TAGCGATTGC CAGTCGCTAT ACAGTCCTCG
251	AAGACTTTG CTCTTATACG TTCTTATCA AAGACAGCGC TTTATGGAGC
301	GACGGAACAC CAATCACCTC CGAAGATATC CGTAACGCTT GGGAGTATGC
351	ACAGGAGAAC TCTCCCCACA TACAGATCTT CCAAGGACTT AACTCTCAA
401	CTCCTTCATC AAATGCAATT ACGATTCTAC TCGACTCGCC CAACCCCGAT
451	TTCCCTAAGC TTCTTGCCCTT TCCTGCATTG GCTATCTTA ACCAGAAAA
501	CCCGAAGCTC TTTAGGGCTC CGTATACTCT TGTAGAGTAT TTCCCAGGGC
551	ATAACATTCA TTTAAAGAAA AACCTTAAC ATTACGACTA CCACTGGCTC
601	TCCATCAACT CCATCAAAC GCTCATTATT CCTGATATAT ATACAGCCAT
651	CCACCTCTA AACAGAGGCA AGGTGGACTG GGTAGGACAA CCCTGGCATC
701	AAGGGATTC TTGGGAGCTC CATAAAACAA CGCAATATCA CTACTACACC
751	TATCCTGTAG AAGGTGCCCTT CTGGCTTTGT CTAATACAA AATCCCCACA
801	CTTAAATGAT CTTCAAAACAA GACATAGACT CGCTACTTGT ATTGATAAAC
851	GTCTCTATCAT TGAAGAAGCT CTTCAAGGAA CCCAACAAACC AGCGGAAACAA
901	CTGTCGGAG GAGCTCCACA ACCAAATCAA TATAAAAAC AAAAGCCTCT
951	AACTCCACAA GAAAAACTCG TGCTTACCTA TCCCTCAGAT ATTCTAAGAT
1001	GCCAACGCAT AGCAGAAATC TTAAAGGAAC AATGGAAAGC TGCTGGAATA
1051	GATTAAATCC TTGAAGGACT CGAAATACCAT CTGTTTGTAA ACAAAACGAAA
1101	AGTCCAAGAC TACGCCATAG CAACACAGAC TGGAGTTGCT TATTACCCAG
1151	GAGCAAATCT AATTCTGAA GAAGACAAGC TCCTGCAAAA CTTTGAGATT
1201	ATCCCGATCT ACTATCTGAG CTATGACTAT CTCACTCAAG ATTTTATAGA
1251	GGGAGTAATC TATAATGCTT CTGGAGCTGT AGATCTCAAA TATACCTATT
1301	TCCCCCTAG

The PSORT algorithm predicts that this protein is an outer membrane lipoprotein (0.790).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 13A.

35 The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 13B) and for FACS analysis. A his-tagged protein was also expressed.

These experiments show that cp6468 is a useful immunogen. These properties are not evident from the sequence alone.

Example 14

40 The following *C.pneumoniae* protein (PID 4376469) was expressed <SEQ ID 27; cp6469>:

1	MKMHLRKPTL KSLIPNLLFL LLTLSSCSKQ KQEPLGKHLV IAMSHDLADL
51	DPRNAYLSRD ASLAKALYEG LTRETDQGIA LALAESYTLS KDHKVYTFKL
101	RPSVWSDGTP LTAYDFEksi KQLYFEEFSP SIHTLLGVIK NSSAIHNAQK
151	SLETLGQAK DDLTLVITLQ QPFPYFLTLI ARPVFSPVHH TLRESYKKGT
201	PPSTYISNGP FVLKKHEHQY YLILEKNPHY YDHESVKLDR VTLKIIIPDAS
251	TATKLFKSks IDWIGSPWSA PISMEDQKVL SQEKILTYSV SSTTLIYNL
301	QKPLIQNKAL RKAIAHAIDR KSILRLVPSG QEAVTLVPPN LSQQLNLQKEI
351	STEEROTKAR AYFQEAKETL SEKELAELSI LYPIDSSNSS IIIAQEIQRQL
401	KDTLGLKIKI QGMeyHCFLK KRRQGDFFIA TGGWIAEVVS PVAFLSILGN
451	PRDLTQWRNS DYEKTLKLY LPHAYKENLK RAEMIIEEET PIIPLYHGKY
501	IYAIHPKIQN TFGSLLGHTD LKNIDILS*

A predicted signal peptide is highlighted.

The cp6469 nucleotide sequence <SEQ ID 28> is:

1 ATGAAGATGC ATAGGCTTAA ACCTACCTTA AAAAGTCTGA TCCCTAATCT
 51 TCTTTCTTA TTGCTCACTC TTTCAAGCTG CTCAAAGCAA AAACAAAGAAC
 101 CCTTAGAAAA ACATCTCGTT ATTGGCATGA GCCATGATCT CGCCGACCTA
 151 GATCCTCGCA ATGCCATTG AAGCAGAGAT GCTTCCCTAG CAAAAGCCCT
 201 CTATGAAGGA CTGACAAGAG AAACGTATCA AGGAATCGCA CTGGCTTGT
 251 CAGAAAGTTA TACCCGTCA AAAGATCATCA AGGTCTATAC CTTTAAACTC
 301 AGACCTTCTG TGTTGAGCGA TGGCACTCCA CTCACGTGTT ATGACTTTGA
 351 AAAATCTATA AAACAACGT ACTTCGAAGA ATTTTCACCT TCCATACATA
 401 CTTTACTCGG CGTGATTAAGG AATTCTTCGG CAATCCACAA TGCTCAAAAA
 451 TCTCTGGAAA CTCTGGGAT ACAGGAAAA GATGATCTA CTTTGGTGT
 501 TACCCCTAGAG CAACCTTCCC CATACTTTCT CACACTTATC GCTCGCCCCG
 551 TATTCTCCCC TGTTCATCAC ACCCTTAGGG AATCCTATAA GAAAGGAACA
 601 CCCCCATCCA CATAACATCTC CAATGGGCC TTTGTCTTAA AAAAACATGA
 651 ACACCAAAAC TACTTAATT TAGAAAAAAA TCCTCACTAC TATGATCATG
 701 AATCAGTAAA GTTAGACCGA GTCACCTTAA AAATTATCCC AGACGCCCTCC
 751 ACAGCCACGA AACTTTCAA AAGTAAATCT ATAGATTGGA TTGGCTCACC
 801 TTGGAGCGCT CCGATATCTA ACGAAGACCA AAAAGTCTC TCCCAAGAAA
 851 AGATTCTTAC CTATTCTGTT TCAAGCACCACCCCTCTTAT CTATAACCTG
 901 CAAAAACCTC TAATACAAAAA TAAAGCCCTC AGGAAAGCCA TTGCTCATGC
 951 TATTGATAGA AAATCTATCT TAAGACTCGT GCCTTCAGGA CAAGAAGCTG
 1001 TAACTCTAGT TCCCCCAAAT CTTTCACAAAC TCAATCTTCA AAAAGAGATC
 1051 TCAACAGAAG AACGACAAAC AAAAGCCAGA GCATATTTC AAGAAGCTAA
 1101 AGAAACACTT TCTGAAAAAG AACTCGCAGA ACTCAGCAGC CTCTATCCTA
 1151 TAGATTCCCTC GAATTCCCTC ATCATAGCTC AAGAAATCCA AAGACAACCTT
 1201 AAAGATAACCT TAGGATTGAA AATCAAAATC CAAGGCATGG AGTACCACTG
 1251 CTTTTTAAAG AAACGTCGTC AAGGAGATTT CTTCATAGCG ACAGGAGGAT
 1301 GGATTGCGGA ATACGTAAGC CCCGTAGCCT TCCTATCTAT TCTAGGCAAC
 1351 CCCAGAGACC TCACACATG GAGAACAGT GATTACGAAA AGACTTTAGA
 1401 GAAACTCTAT CTCCCTCATG CCTACAAAGA GAATTAAAAA CGCGCAGAAA
 1451 TGATAATAGA AGAAGAAACC CCGGATTATCC CCCTGTATCA CGGCAAATAT
 1501 ATTACGCTA TACATCTAA AATCCAGAAT ACATTCGGAT CTCTTCTAGG
 1551 CCACACAGAT CTCAAAATA TCGATATCTT AAGTTAG

The PSORT algorithm predicts a periplasmic location (0.934).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 14A.

35 The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 14B) and for FACS analysis. A his-tagged protein was also expressed.

These experiments show that cp6469 is a useful immunogen. These properties are not evident from the sequence alone.

Example 15

40 The following *C.pneumoniae* protein (PID 4376602) was expressed <SEQ ID 29; cp6602>:

1 MAASGGTGGL GGTQGVNLAA VEAAAAKADA AEVVVASQEGS EMNMIQQSQD
 51 LTNPAAATRT KKKEEKFQTL ESRKKGEAGK AEKKSESTEE KPDTDLADKY
 101 ASGNSEISGQ ELRGLRDAIG DDASPEDILA LVQEIKDPA LQSTALDYLV
 151 QTPPPSQGKL KEALIQARNT HTEQFGRTAI GAKNILFASQ EYADQLNVSP
 201 SGLRSLYLEV TGDTHTCDQL LSMIQLDRYTY QDMAIVSSFL MKGMATELKR
 251 QGPYVPSAQL QVLMTETRNL QAVLTSYDF ESRVPILLDS LKAEGIQTPS
 301 DLNFVKVAES YHKIINDKFP TASKVEREVR NLIGDDVDSS TGVLNLFFSA
 351 LRQTSRSLFS SADKRQQLGA MIANALDAVN INNEDYPKAS DFPKPYPWS*

The cp6602 nucleotide sequence <SEQ ID 30> is:

50 1 ATGGCAGCAT CAGGAGGCAC AGGTGGTTA GGAGGCACTC AGGGTGTCAA
 51 CCTTGCAGCT GTAGAAGCTG CAGCTGCAAA AGCAGATGCA GCAGAAGTTG
 101 TAGCCAGCCA AGAAGGTTCT GAGATGAACA TGATTCAACA ATCTCAGGAC
 151 CTGACAAATC CCGCAGCAGC AACACGCACG AAAAAGGAGG AAGAGAAGTT
 201 TCAAACCTCA GAATCTCGGA AAAAAGGAGA AGCTGGAAG GCTGAGAAAAA
 251 AATCTGAATC TACAGAAGAG AAGCCTGACA CAGATCTTGC TGATAAGTAT
 301 GCTTCTGGGA ATTCTGAAT CTCTGGTCAA GAACTTCGGC GCCTGCGTGA
 351 TGCAATAGGA GACGATGCTT CTCCAGAAGA CATTCTTGCT CTTGTACAAG

401 AGAAAATTAA AGACCCAGCT CTGCAATCCA CAGCTTGGA CTACCTGGTT
 451 CAAACGACTC CACCCTCCC AGGTAAATTAA AAAGAACGCG TTATCCAAGC
 501 AAGGAATACT CATA CGGAGC AATT CGGACG AACTGCTATT GGTGCGAAAAA
 551 ACATCTTATT TGCCCTCTCAA GAATATGCAG ACCAACTGAA TGTTTCTCCT
 601 TCAGGGCTTC GCTCTTGTA CTTAGAAGTG ACTGGAGACA CACATACCTG
 651 TGATCAGCTA CTTTCTATGC TTCAAGACCG CTATACCTAC CAAGATATGG
 701 CTATTGTCAG CTCCCTTCTA ATGAAAGGA TGGCAACAGA ATTAAAAAGG
 751 CAGGGTCCCT ACGTACCCAG TGCGCAACTA CAAGTTCTCA TGACAGAAAC
 801 TCGTAACCTG CAAGCAGTTC TTACCTCGTA CGATTACTTT GAAAGTCGCG
 851 TTCCCTATTTT ACTCGATAGC TTAAAAGCTG AGGGAACTCCA AACTCCTTCT
 901 GATCTAAACT TTGTGAAGGT AGCTGAGTCC TACCATAAAA TCATTAACGA
 951 TAAGTTCCCA ACAGCATCTA AAGTAGAACG AGAAGTCGCG AATCTCATAG
 1001 GAGACGATGT TGATTCTGTG ACCGGTGTCT TGAACCTATT CTTTTCTGCT
 1051 TTACGTCAAA CGTCGTACG CCTTTTCTCT TCAGCAGACA AACGTCAGCA
 1101 ATTAGGAGCT ATGATTGCTA ATGCTTTAGA TGCTGTAAAT ATAACAAATG
 1151 AAGATTATCC CAAAGCATCA GACTTCCCTA AACCTATCC TTGGTCATGA

The PSORT algorithm predicts a cytoplasmic location (0.080).

The protein was expressed in *E.coli* and purified as both a His-tag and a GST-fusion product, as shown in Figure 15A. The recombinant proteins were used to immunise mice, whose sera were used 20 in a Western blot (Figure 15B) and for FACS analysis (Figure 15C).

The cp6602 protein was also identified in the 2D-PAGE experiment (Cpn0324).

These experiments show that cp6602 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 16

25 The following *C.pneumoniae* protein (PID 4376727) was expressed <SEQ ID 31; cp6727>:

1 MKYSLPWLLT SSALVFSLHP LMAANTDLSS SDNYENGSSG SAAFTAKETS
 51 DASGTTTYTLT SDVSITNVSA ITPADKSCFT NTGGALSFVG ADHSLVLQTI
 101 ALTHDGAAIN NTNTALSFSG FSSLLIDSAP ATGTSGGKGA ICVTNTEGGT
 151 ATFTDNASVT LQKNTSEKDG AAVSAYSIDL AKTTTAALLD QNTSTKNGGA
 201 LCSTANTTVQ GNNSGTVTFSS NTATDKGGGI YSKEDDSLTD ANTGVVTFKS
 251 NTAKTGGAWS SDDNLALTGN TQVLFQENKT TGSAAQANNP EGCGBAACCY
 301 LATATDKTGL AISQNQEMSF TSNTTTTANGG AYAITKCTLD GNTTLTFDQN
 351 TATAGCGGAI YTTETEDFLSK GSTGTVTFST NTAKTGGALY SKGNSSLTGN
 401 TNLLFSGNKA TGPSNSSANQ ECGGGAILAF IDSGSVSDKT GLSIANNQEV
 451 SLTSNAATVS GGAIYATKCT LTGNGNSLTFD GNTAGTSGGA IYTETEDFTL
 501 TGSTGTVTF SNTAKTGGAL YSKGNNSLSG NTNLFSGNK ATGPSNSSAN
 551 QEGCGGAILS FLESASVSTK KGLWIEDNEN VSLSGNTATV SGAIYATKC
 601 ALHGNTTLTF DGNTAETAGG AIYTETEDFT LTGSTGTVTF STNTAKTAGA
 651 LHTKGNTSFT KNKALVFSGN SATATATTTT DQEKGCGAIL CNISESDIAT
 701 KSLTLTENES LSFINNTAKR SGGGIYAPKC VISGSESINF DGNTAETSGG
 751 AIYSKNLISI ANGPV SFTNN SGGKGGAIYI ADSGELSLEA IDGDITFSGN
 801 RATEGTSTPN SIHLGAGAKI TKLAAAPGHT IYFYDPITME APASGGTIEE
 851 LVINPVVKAI VPPPQPKNGP IASVPVVPVA PANPNTGTIV FSSGKLPSQD
 901 ASIPANTTTI LNQKINLAGG NVVKEGATL QVYSFTQYPD STVFM DAGTT
 951 LETTTTNNNT GSIDLKNLNV NLDA LDGKRM ITIAVNSTSG GLKISGDLKF
 1001 HNNEGSFYDN PGLKANLNLP FLDLSSTS GT VNLDDFNPPIP SSMAAPDYGY
 1051 QGSWTLVPKV GAGGKVTLVA EWQALGYTPK PELRATLVPN SLWNAYVNIH
 1101 SIQOEIATAM SDAPSHPGIW IGGIGNAFHQ DKQKENAGFR LISRGYIVGG
 1151 SMITTPQEYTF AVAFSQLFGK SKDYVVSDIK SQVYAGSLCA QSSYVIPLHS
 1201 SLRRHVLSKV LPELPGETPL VLHGQVSYGR NHNNMTTKLA NNTQGKSDWD
 1251 SHSFAVEVGG SLPVDLNRYR LTSYSPYVKL QVVSVNQKGF QEVAADPRIF
 1301 DASHLVNVSI PMGLTFKHES AKPPSALLLT LGYAVDAYRD HPHCLTSLTN
 1351 GTSWSTFATN LSRQAFFAEA SGHLKLLHGL DCFASGSCEL RSSSRSYNAN
 1401 CGTRYSF*

55 A predicted signal peptide is highlighted.

The cp6727 nucleotide sequence <SEQ ID 32> is:

	1	ATGAAATATT CTTTACCTTG GCTACATTACC TCTTCGGCTT TAGTTTTCTC
	51	CCTACATCCA CTAATGGCTG CTAACACGGA TCTCTCATCA TCCGATAACT
5	101	ATGAAAATGG TAGTAGTGGT AGCCGAGCAT TCACTGCCAA GGAAACTTCG
	151	GATGCTTCAG GAACTACCTA CACTCTCACT AGCGATGTT CTATTACGAA
	201	TGTATCTGCA ATTACTCTG CAGATAAAAG CTGTTTACA AACACAGGAG
	251	GAGCATTGAG TTTTGTGGA GCTGATCACT CATTGGTTCT GCAAACCATA
10	301	GCGCTTACGC ATGATGGTC TGCAATTAAAC AATACCAACA CAGCTTTTC
	351	TTTCTCAGGA TTCTCGTCAC TCTTAACTCGA CTCAGCTCCA GCAACAGGAA
15	401	CTTCGGGGCG CAAGGGTGCT ATTGTGTGA CAAAATACAGA GGGAGGTTACT
	451	GCGACTTTA CTGACAAATGC CAGTGTCACT CTCCAAAAAA ATACTTCAGA
	501	AAAAGATGGA GCTGCAGTTT CTGCCTACAG CATCGATCTT GCTAAAGACTA
	551	CGACAGCGC TCTCTTAGAT CAAAATACTA GCACAAAAAA TGGCGGGGCC
20	601	CTCTGTAGTA CAGCAAACAC TACAGTCCAA GGAAACTCAG GAACGGTGAC
	651	CTTCTCCTCA AATACTGCTA CAGATAAAAGG TGGGGGGATC TACTCAAAG
	701	AAAAGGATAG CACGCTAGAT GCCAATACAG GAGTCGTTAC CTTCAAATCT
	751	AATACTGCAA AGACGGGGGG TGCTTGGAGC TCTGATGACA ATCTTGCTCT
	801	TACCGGCAAC ACTCAAGTAC TTTTCAGGA AAATAAAACA ACCGGCTCAG
25	851	CAGCACAGG AAATAACCCG GAAGGTTGTG GTGGGGCAAT CTGTTGTTAT
	901	CTTGCTACAG CAACAGACAA AACTGGATTA GCCATTTCTC AGAATCAAGA
	951	AATGAGCTTC ACTAGTAATA CAACAACCTGC GAATGGTGGA GCGATCTACG
	1001	CTACTAAATG TACTCTGGAT GGAAACACAA CTCTTACCTT CGATCAGAAT
	1051	ACTGCGACAG CAGGATGTGG CGGAGCTATC TATACAGAAA CTGAAGATT
30	1101	TTCTCTTAAAG GGAAGTACGG GAACCGTGAC CTTCAGCACA AATACAGCAA
	1151	AGACAGGGGG CGCCTTATAT TCTAAAGGAA ACAGCTCGT GACTGGAAAT
	1201	ACCAACCTGC TCTTTTCAGG GAACAAAGCT ACGGGGCCGA GTAATTCTTC
	1251	AGCAAATCAA GAGGGTTGCG GTGGGGCAAT CCTAGCCTT ATTGATTCA
	1301	GATCCGTAAG CGATAAAACA GGACTATCGA TTGCAAACAA CCAAGAACGTC
	1351	AGCCTCACTA GTAATGCTGC AACAGTAAGT GGTGGTGGA TCTATGCTAC
35	1401	CAAATGTACT CTAACTGGAA ACGGCTCCCT GACCTTGAC GGCAATACTG
	1451	CTGGAACCTTC AGGAGGGCG ATCTATACAG AACTGAAGA TTTTACTCTT
	1501	ACAGGAAGTA CAGGAACCGT GACCTTCAGC ACAAAATACAG CAAAGACAGG
	1551	CGGCGCCTTA TATTCTAAAG GCAACAACCTC TCTGTCTGT AATACCAACC
	1601	TGCTCTTTTC AGGGAAACAAA GCTACGGGCC CGAGTAATTC TTCAGCAAAT
40	1651	CAAGAGGGTT GCGGTGGGC AATCCTATCG TTTCTTGAGT CAGCATCTGT
	1701	AAGTACTAAA AAAGGACTCT GGATTGAAGA TAACGAAAAC GTGAGTCTCT
	1751	CTGGTAATAC TGCAACAGTA AGTGGCGGTG CGATCTATGC GACCAAGTGT
	1801	GCTCTGCATG GAAACACGAC TCTTACCTT GATGGCAATA CTGCCGAAAC
	1851	TGCAAGGAGGA GCGATCTTC CAGAAACCGA AGATTTACT CTTACGGGAA
45	1901	GTACGGGAAC CGTGACCTTC AGCACAAATA CAGCAAAGAC AGCAGGGCT
	1951	CTACACTA AAGGAAATAC TTCTTTACC AAAAATAAGG CTCTTGATT
	2001	TTCTGGGAAAT TCAGCAACAG CAACAGCAAC AACAACCTACA GATCAAGAAG
	2051	GTGTGGTGG AGCGATCTC TGTAATATCT CAGAGTCTGA CATAGCTACA
	2101	AAAAGCTTAA CTCTTACTGA AAATGAGAGT TTAAGTTCA TTAACAATAC
50	2151	GGCAAAAAGA AGTGGTGGTG GTATTTATGC TCCTAACTGT GTAATCTCAG
	2201	GCAGTGAATC CATAAAACTTT GATGGCAATA CTGCTGAAAC TTCTGGAGGA
	2251	GGGATTTATT CGAAAACCT TTGATTACA GCTAACGGTC CTGCTCTCTT
	2301	TACCAATAAT TCTGGAGGCA AGGGAGGCGC CATTATATA GCCGATAGCG
	2351	GAGAACTTTC CTTAGAGGCT ATTGATGGGG ATATTACTTT CTCAGGGAAC
55	2401	CGAGCGACTG AGGAAACCTTC AACTCCCCAAC TCGATCCATT TAGGTGCAGG
	2451	GGCTAAGATC ACTAAGCTTG CAGCAGCTCC TGGTCATACG ATTTATTTT
	2501	ATGATCCTAT TACGATGGAA GCTCCTGCAT CTGGAGGAAC AATAGAGGAG
	2551	TTAGTCATCA ATCCTGTGT CAAAGCTATT GTTCCCTCTC CCCAACAAA
	2601	AAATGGTCCT ATAGCTCAG TGCCTGTAGT CCCTGTAGCA CCTGCAAACC
60	2651	CAAACACGGG AACTATAGTA TTTCTTCTG GAAAACCTCC CAGTCAGAAT
	2701	GCCTCGATTC CTGCAAATAC TACCAACATA CTGAACCGAGA AGATCAAAC
	2751	AGCAGGAGGA AATGTCGTTT TAAAAGAAGG AGCCACCCCTA CAAGTATATT
	2801	CCTTCACACA GCAGCCTGAT TCTACAGTAT TCATGGATGC AGGAACGACC
	2851	TTAGAGACCA CGACAACCAA CAATACAGAT GGCAGCATCG ATCTAAAGAA
65	2901	TCTCTCTGTA AATCTGGATG CTTTAGATGG CAAGCGTATG ATAACGATTG
	2951	CCGTAAACAG CACAAGTGGG GGATTTAAAAA TCTCAGGGGA TCTGAAATT
	3001	CATAACAAATG AAGGAAGTTT CTATGACAAT CCTGGGTGAA AAGCAAAC
	3051	AAATCTTCTC TTCTTAGATC TTTCTTCTAC TTCAGGAAC GTAAATTTAG
	3101	ACGACTTCAGA TCCGATTCCT TCTAGCATGG CTGCTCCGGAA TTATGGGTAT
	3151	CAAGGGAGTT GGACTCTGGT TCCTAAAGTA GGAGCTGGAG GGAAGGTGAC
	3201	TTTGGTCGCG GAATGGCAAG CGTTAGGATA CACTCCTAAA CCAGAGCTTC
	3251	GTGCGACTTT AGTTCTTAAT AGCCTTGGGAA ATGCTTATGT AAACATCCAT

5 3301 TCTATACAGC AGGAGATCGC CACTGCATG TCGGACGCTC CCTCACATCC
 3351 AGGGATTGAG ATTGGAGGTA TTGGCAACGC CTTCCATCAA GACAAGCAAA
 3401 AGGAAAATGAG AGGATTCCGT TTGATTCCA GAGGTATATA TGTTGGTGGC
 3451 AGCATGACCA CCCCTCAAGA ATATACCTT GCTGTTGCAT TCAGCCAATC
 3501 CTTTGGCAAA TCTAAGGATT ACGTAGTC CCGATATTAAA TCTCAAGTCT
 3551 ATGCAGGATC TCTCTGTGCT CAGAGCTCTT ATGTCATTCC CCTGCATAGC
 3601 TCATTACGTC GCCACGTCT CTCTAAGGTC CTTCCAGAGC TCCCAGGAGA
 3651 AACTCCCCCTT GTTCTCCATG GTCAAGTTTC CTATGGAAGA AACCAACATA
 3701 ATATGACGAC AAAGCTTGCG AACAACACAC AAGGGAAATC AGACTGGGAC
 3751 AGCCATAGCT TCGCTGTGA AGTCGGTGGT TCTCTTCTG TAGATCTAAA
 3801 CTACAGATAC CTTACCAAGCT ACTCTCCCTA TGTGAAACTC CAAGTTGTGA
 3851 GTGTAATCA AAAAGGATTC CAAGAGGTTG CTGCTGATCC ACGTATCTT
 3901 GACGCTAGCC ATCTGGTCAA CGTGTCTATC CCTATGGGAC TCACCTTCAA
 3951 ACACGAATCA GCAAAAGCCCC CCAGTGTCTT GCTTCTTACT TTAGGTTACG
 4001 CTGTAGATGC TTACCGGGAT CACCCCTCACT GCCTGACCTC CTTAACAAAT
 4051 GGCACCTCGT GGTCTACGTT TGCTACAAAC TTATCACGAC AAGCTTCTT
 4101 TGCTGAGGCT TCTGGACATC TGAAGTTACT TCATGGTCTT GACTGCTTCG
 4151 CTTCTGGAAG TTGTGAACGT CGCAGTCCT CAAGAAGCTA TAATGCAAAC
 4201 TGTGGAACTC GTTATTCTTT CTAA

20 The PSORT algorithm predicts an outer membrane location (0.915).

The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 16A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 16B) and for FACS analysis (Figure 16C). A GST-fusion protein was also expressed.

The cp6727 protein was also identified in the 2D-PAGE experiment (Cpn0444).

25 These experiments show that cp6727 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 17

The following *C.pneumoniae* protein (PID 4376731) was expressed <SEQ ID 33; cp6731>:

30 1 MKSSLHWFLI SSSLALPLSL NFSAAVVE INLGPTNSFS GPGTYTPPAQ
 51 TTNADGTYIN LTGDVSITNA GSPTALTASC FKETTGTLF QGHGYQFLQQ
 101 NIDAGANCTF TNTAANKLLS FSGFSYLSLI QTNNATTGTG AIKSTGACSI
 151 QSNYSCYFGQ NFSNDNGGAL QGSISLSLN PNLTFAKNKA TQKGALYST
 201 GGITINNTLN SASFSENAA NNGGAIYTEA SSFISSNKA SFINNSVTAT
 251 SATGGAIYCS STSAPKPVLT LSDNGELNFI GNTAITSGGA IYTDNLVLSS
 301 GGPFLFKNNNS AIDTAAPLGG AIAIADSGSL SLSALGGDIT FEGNTVVKG
 351 SSSQTTRNS INIGNTNAKI VQLRASQGNT IIFYDPITTS ITAALSDALN
 401 LNGPDLAGNP AYQGTIVFSC EKLSEAAEADNLKSTIQQ PLTIAGGQLS
 451 LKSGVTLVAK SFSQSPGSTL LMDAGTTLET ADGITINNLV LNVDSLKETK
 501 KATLKATQAS QTVTLSGSL LVDPSGNVYE DVSWNNPQVF SCLTLTADDP
 551 ANIHITDLAA DPLEKNPPIHW GYQGNWALSW QEDTATKSATLWTWKTGY
 601 NPNPERRGTL VANTLWGSFV DVRSIQQLVA TKVRQSQETR GIWCEGISNF
 651 FHKDSTKINK GFRHISAGYV VGATTTLASD NLITAACFQL FGKDRDHFIN
 701 KNRASAYAAS LHLQHLATLS SPSLLRYLPG SESEQPVLFD AQISYTIYSKN
 751 TMKTYYTQAP KGESSWYNDG CALELASSLP HTALSHEGLF HAYFPPIKVE
 801 ASYIHQDSFK ERNTTLVRSF DSGDLINSV PIGITFERFS RNERASYEAT
 851 VIYVADVYRK NPDCTTALLI NNNTSWKTTGT NLSRQAGIGR AGIFYAFSPN
 901 LEVTSNLSME IRGSSRSYNA DLGGKFQF*

A predicted signal peptide is highlighted.

The cp6731 nucleotide sequence <SEQ ID 34> is:

50 1 ATGAAATCCT CTCTTCATTG GTTTTTAATC TCGTCATCTT TAGCACTTCC
 51 CTTGTCACTA AATTTCCTCTG CGTTGCTGC TGTTGTTGAA ATCAATCTAG
 101 GACCTACCAA TAGCTTCTCT GGACCAGGAA CCTACACTCC TCCAGCCCAA
 151 ACAACAAATG CAGATGGAAC TATCTATAAT CTAACAGGGG ATGTCCTCAAT
 201 CACCAATGCA GGATCTCCGA CAGCTCTAAC CGCTTCCTGC TTTAAAGAAA

	251	CTACTGGGAA	TCTTTCTTTC	CAAGGCCACG	GCTACCAATT	TCTCCCTACAA
5	301	AATATCGATG	CGGGAGCGAA	CTGTACCTTT	ACCAATAACAG	CTGCAAATAA
	351	GCTTCTCTCC	TTTCAGGAT	TCTCCTATTT	GTCACTAATA	CAAACCACGA
	401	ATGCTTACAC	AGGAACAGGA	GCCATCAAGT	CCACAGGAGC	TTGTTCTATT
	451	CAGTCGAACT	ATAGTTGCTA	CTTTGGCCAA	AACTTTCTA	ATGACAATGG
10	501	AGGCGCCCTC	CAAGGCAGCT	CTATCAGTCT	ATCGCTAAC	CCCAACCTAA
	551	CGTTGCCAA	AAACAAAGCA	ACGCCAAAAG	GGGGTGCCCT	CTATTCCACG
	601	GGAGGGATTA	CAATTAACAA	TACGTTAAAC	TCAGCATCAT	TTTCTGAAAAA
15	651	TACCGCGGCG	AAACATGGCG	GAGCCATTAA	CACGGAAGCT	AGCAGTTTTA
	701	TTAGCAGCAA	CAAAGCAATT	AGCTTTATAA	ACAATAGTGT	GACCGCAACC
	751	TCAGCTACAG	GGGGAGCCAT	TTACTGTAGT	AGTACATCAG	CCCCCAAACC
	801	AGTCTTAACT	CTATCAGACA	ACGGGAACT	GAACCTTATA	GGAAATACAG
	851	CAATTACTAG	TGGTGGGGCG	ATTTATACTG	ACAATCTAGT	TCTTTCTCT
20	901	GGAGGACCTA	CGCTTTTAA	AAACAACCTCT	GCTATAGATA	CTGCAGCTCC
	951	CTAGGAGGA	GCAATTGCGA	TTGCTGACTC	TGGATCTTIG	AGTCTTTCGG
	1001	CTCTGGTGG	AGACATCACT	TTGAAGGAA	ACACAGTAGT	CAAAGGAGCT
	1051	TCTTCGAGTC	AGACCACACTAC	CAGAAATTCT	ATTAACATCG	GAAACACCAA
	1101	TGCTAAGATT	GTACAGCTGC	GAGCCTCTCA	AGGCAAAACT	ATCTACTTCT
	1151	ATGATCCTAT	AAACAACTAGC	ATCACTGCAG	CTCTCTCAGA	TGCTCTAAC
25	1201	TTAAATGGTC	CTGACCTTGC	AGGGAATCCT	GCATATCAAG	GAACCATCGT
	1251	ATTTTCTGGG	GAGAAGCTCT	CGGAAGCAGA	AGCTGCAGAA	GCTGATAATC
	1301	TCAAATCTAC	AATTCAAGCAA	CCTCTAACTC	TTGCGGGAGG	GCAACTCTCT
	1351	CTTAAATCAG	GAGTCACCTCT	AGTGCTAAG	TCCCTTTCGC	AATCTCCGGG
	1401	CTCTACCCCTC	CTCATGGATG	CAGGGACCAC	ATTAGAAACC	GCTGATGGGA
	1451	TCACTATCAA	TAATCTTGT	CTCAATGTAG	ATTCCTTAAA	AGAGACCAAG
	1501	AAGGCTACGC	AAAAAGCAAC	ACAAGCAAGT	CAGACAGTCA	CTTTATCTGG
	1551	ATCGCTCTC	CTTGCTAGATC	CTTCTGGAAA	TGTCTACGAA	GATGTCCTTT
30	1601	GGAAATAACCC	TCAAGTCTT	TCTGTCTCA	CTCTTACTGC	TGACGACCCC
	1651	GCGAATATTTC	ACATCACAGA	CTTAGCTGCT	GATCCCCTAG	AAAAAAATCC
	1701	TATCCATTGG	GGATACCAAG	GGATTGGGC	ATTATCTTGG	CAAGAGGATA
	1751	CTGCGACTAA	ATCCAAGCA	GCGACTCTTA	CCTGGACAAA	AACAGGATAC
	1801	AATCCGAATC	CTGAGCGTCG	TGGAACCTTA	GTTGCTAAC	CGCTATGGGG
	1851	ATCCTTTGTT	GATGTGCGCT	CCATACAACA	GCTTGTAGCC	ACTAAAGTAC
35	1901	GCCAATCTCA	AGAAAACCTGC	GGCATCTGGT	GTGAAGGGAT	CTCGAACITC
	1951	TTC CATAAAAG	ATAGCAGCAA	GATAAAATAA	GGTTTTCGCC	ACATAAGTGC
	2001	AGGTTATGTT	GTAGGAGCGA	CTACAACATT	AGCTTCTGAT	AATCTTATCA
	2051	CTGCAGCCCT	CTGCCAATT	TTGGGAAAG	ATAGAGATCA	CTTTATAAAT
	2101	AAAAATAGAG	CTTCTGCCA	TGCAGCTCT	CTCCATCTCC	AGCATCTAGC
40	2151	GACCTTGTCT	TCTCCAAGCT	TGTTACGCTA	CCTTCCTGG	TCTGAAAGTG
	2201	AGCAGCCTGT	CCTCTTTGAT	GTCAGATCA	GCTATATCTA	TAGTAAAAAT
	2251	ACTATGAAAA	CCTATTACAC	CCAAGCACCA	AAGGGAGAGA	GCTCGTGGTA
	2301	TAATGACGGT	TGCGCTCTGG	AACTTGCAG	CTCCCTACCA	CACACTGCTT
	2351	TAAGCCATGA	GGGTCTCTTC	CACCGTATT	TTCTTTCTAT	CAAAGTAGAA
45	2401	GCTCTCTAAC	AAACATAACCT	CGTGGAAAAC	TACAGGAACG	AATCTCTCAA
	2451	ACGATCTTTC	GATAGCGGTG	ATTTAATTAA	CGTCTCTGTG	CCTATTGGAA
	2501	TTACCTTCGA	GAGATTCTCG	AGAAACGAGC	GTGCGTCTTA	CGAAGCTACT
	2551	GTCATCTACG	TTGCCGATGT	CTATCGTAAG	AATCCTGACT	GCACGACAGC
	2601	TCTCTAAC	AAACATAACCT	CGTGGAAAAC	TACAGGAACG	AATCTCTCAA
50	2651	GACAAGCTGG	TATCGGAAGA	GCAGGGATCT	TTTATGCCCT	CTCTCCAAAT
	2701	CTTGAGGTCA	CAAGTAACCT	ATCTATGGAA	ATTCTGTGGAT	CTTCACGCAG
	2751	CTACAATGCA	GATCTGGAG	GTAAGTTCCA	GTTCTAA	

The PSORT algorithm predicts an outer membrane location (0.926).

The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 17A. A GST-fusion protein was also expressed. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 17B; his-tag) and for FACS analysis (Figure 17C; his-tag and GST-fusion).

The GST-fusion protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis. Less cross-reactivity was seen with the his-fusion.

These experiments show that cp6731 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 18

The following *C.pneumoniae* protein (PID 4376737) was expressed <SEQ ID 35; cp6737>:

5	1	MPLSFKSSSE	CLLAICLCSAS	CAFAETRLGG	NFVPPITNQG	EEILLTSDFV
	51	CSNFLGASFS	SSFINSSSNL	SLIGKGLSLT	FTSCQAPTN	NYALLSAAET
	101	LTFKNFSSIN	FTGNQSTGLG	GLIYGKDIVF	QSIKDLIFTT	NRVAYSPASV
	151	TTSATPAITT	VITGASALQP	TDSLTVENIS	QSIKFFGMIA	NFGSAISSSP
10	201	TAVVKFINNT	ATMSFSHNFT	SSGGVIYGG	SSLLFENMSG	CIIFTANSVC
	251	NSLKGVTTPSS	GTYALGSGGA	ICIPTGTTEL	KNNQGKCTFS	YNGTPNDAGA
	301	IYAETCNIVG	NQGALLLDSN	TAARNGGAIC	AKVLNIQGRG	PIEFSRNRAE
	351	KGGAIFIGPS	VGDPAKQTST	LTLASEGDI	AFQGNMLNTK	PGIRNAITVE
	401	AGGEIVSLSA	QGGSRLVFYD	PITHSLPTTS	PSNKDITINA	NGASGSVVFT
15	451	SKGLSSTELL	LPANTTTILL	GTVKIASGEL	KITDNAVVNV	LGFATQGSGQ
	501	LTLGSGGTLG	LATPTGAPAA	VDFTIGKLA	DPPFSFLKRDF	VSASVNAGTK
	551	NVTLTGALVL	DEHDVTDLYD	MVSLQTPVAI	PIAVFKGATV	TKTGFDPGEI
	601	ATPPSHYGYQG	KWSYTWSRPL	LIPAPDGFP	GGPSPSANTL	YAVWNSDTLV
20	651	RSTYILDPER	YGEIVSNSLW	ISFLGNQAFS	DILQDVLLID	HPLSITAKA
	701	LGAYVEHTPR	QHEGFSGRY	GGYQAALSMN	YTDHTTLGLS	FGQLYGKTNA
	751	NPYDSRCSEQ	MYLLSFFGQF	PIVTQKSEAL	ISWKAAYGYS	KNHLNTTYLR
	801	PDKAPKSQGQ	WHNNSYVLI	SAEPFFLNW	LLTRPLAQAW	DLSGFISAER
	851	LGGWQSKFTE	TGDLQRFSR	GKGVNVSLPI	GCSSQWFTP	KKAPSTLTIK
	901	LAYKPDYRV	NPHNIVTVVS	NQUESTSISGA	NLRRHGLFVQ	IHDVVDLTED
	951	TQ AFLNYTFD	GKNGFTNHVR	STGLKSTF*		

25 A predicted signal peptide is highlighted.

The cp6737 nucleotide sequence <SEQ ID 36> is:

30	1	ATGCCTCTTT	CTTTCAAATC	TTCATCTTT	TGTCTACTTG	CCTGTTATG
	51	TAGTCAAGT	TGGCGTTTG	CTGAGACTAG	ACTCGGAGGG	AACTTTGTT
	101	CTCCAATTAC	GAATCAGGGT	GAAGAGATCT	TACTCACTTC	AGATTTGTT
	151	TGTTCAAACT	TCTTGGGGC	GAGTTTTCA	AGTTCTTTA	TCAATAGTT
	201	CAGCAATCTC	TCCCTTATTAG	GGAAAGGCC	TTCTTAAACG	TTTACCTCTT
	251	GTCAAGCTCC	TACAAATAGT	AACTATGCGC	TACTTTCTGC	CGCAGAGACT
	301	CTGACCTTCA	AGAATTTC	TTCTATAAAC	TTTACAGGG	ACCAATCGAC
	351	AGGACTTGGC	GGCCTCATCT	ACGGAAAAGA	TATTGTTTC	CAATCTATCA
35	401	AAGATTTGAT	CTTCACTACG	AACC GTTG	CCTATTCTCC	AGCATTCTGTA
	451	ACTACGTCGG	CAACTCCCGC	AATCACTACA	GTAAC TACAG	GAGCCTCTGC
	501	TCTCCAACCT	ACAGACTCAC	TCACTGTCGA	AAACATATCC	CAATCGATCA
	551	AGTTTTTGG	GAACCTTGCC	AACTTCGGCT	CTGCAATTAG	CAGTTCTCCC
40	601	ACGGCAGTCG	TTAAATTCA	CAATAACACC	GCTACCATGA	GCTTCTCCC
	651	TAACTTTACT	TCGTCAGGAG	GCGCGTGT	TTATGGAGGA	AGCTCTCTCC
	701	TTTTGAAAAA	CAATTCTGGA	TGCATCATCT	TCACCGCAA	CTCCGTGTG
	751	AACAGCTTAA	AAGGCGTCAC	CCCTTCATCA	GGAACCTATG	CTTTAGGAAG
	801	TGGCGGAGCC	ATCTGCATCC	CTACGGGAAC	TTTCAATTAA	AAAACAATC
45	851	AGGGGAAGTG	CACCTCTCT	TATAATGGTA	CACCAAATGA	TGCGGGTGC
	901	ATCTACGCCG	AAACCTGCAA	CATCGTAGGG	AACCAGGG	CCTTGCTCCT
	951	AGATAGCAAC	ACTGCAGCGA	GAATGGCGG	AGCCATCTGT	GCTAAAGTGC
	1001	TCAAATATTCA	AGGACGCGGT	CCTATTGAA	TCTCTAGAAA	CCGCGCGGAG
	1051	AAGGGTGGAG	CTATTTTCAT	AGGCCCTCT	GTTGGAGACC	CTGCGAAGCA
50	1101	AACATCGACA	CTTACGATT	TGGCTTCCGA	AGGTGATATT	CGCGTCCAAG
	1151	GAAACATGCT	CAATACAAAAA	CCTGGAATCC	GCAATGCCAT	CACTGTAGAA
	1201	GCAGGGGGAG	AGATTCGTG	TCTATCTGCA	CAAGGAGGCT	CACGTCTTGT
	1251	ATTTTATGAT	CCCATTACAC	ATAGCCTCCC	AACCACAAGT	CCGTCTAATA
	1301	AAGACATTAC	AATCAACGCT	AA TGGCGCTT	CAGGATCTGT	AGTCTTTACA
	1351	AGTAAGGGAC	TCTCCTCTAC	AGA ACTCCTG	TTGCGCTGCCA	ACACGACAAC
	1401	TATACTTCTA	GGAACAGTC	AGATCGCTAG	TGGAGAACTG	AAGATTACTG
	1451	ACAATGCGGT	TGTCAATGTT	CTTGGCTTCG	CTACTCAGGG	CTCAGGTCA
55	1501	CTTACCCCTGG	GCTCTGGAGG	AA CCTT TAGGG	CTGGCAACAC	CCACGGGAGC
	1551	ACCTGCGCT	GTAGACTTTA	CGATTGGAAA	GTTAGCATTC	GATCCTTTTT
	1601	CCTTCCTAAA	AAGAGATTT	GTTTCAGCAT	CAGTAAATGC	AGGCACAAAAA
60	1651	AACGTCACTT	TAACAGGAGC	TCTGGTTCTT	GATGAACATG	ACGTTACAGA

1701 TCTTTATGAT ATGGTGTCA TACAAACTCC AGTAGCAATT CCTATCGCTG
 1751 TTTTCAAAGG AGCAACCGTT ACTAAGACAG GATTCCTGA TGGGGAGATT
 1801 GCGACTCCAA GCCACTACGG CTACCAAGGA AAGTGGTCCT ACACATGGTC
 1851 CCGTCCCCCTG TTAATTCCAG CTCCGTATGG AGGATTTCTC GGAGGTCCCT
 1901 CTCCCTAGCGC AAATACTCTC TATGCTGTAT GGAATTCAAGA CACTCTCGTG
 1951 CGTTCTACCT ATATCTTAGA TCCCAGCGT TACGGAGAAA TTGTCAGCAA
 2001 CAGCTTATGG ATTTCCTCT TAGGAATCA CGCATTCTCT GATATTCTCC
 2051 AAGATGTTCT TTTGATAGAT CATCCCAGGT TGTCATAAC CGCGAAAGCT
 2101 TAGGAGCCT ATGTCGAACA CACACCAAGA CAAGGACATG AGGGCTTTTC
 2151 AGGTCGCTAT GGAGGCTACC AAGCTGCGCT ATCTATGAAC TACACGGACC
 2201 ACAACTACGTT AGGACTTCTT TTCCGGCAGC TTTATGGAAA AACTAACGCC
 2251 AACCCCTACG ATTACACGTTG CTCAGAACAA ATGTATTTCAC TCTCGTTCTT
 2301 TGGTCAATTG CCTATCGTG CTCAAAAGAG CGAGGCCTTA ATTCCTGGAA
 2351 AAGCAGCTTA TGGTTATTCC AAAAATCACC TAAATACAC CTACCTCAGA
 2401 CCTGACAAAG CTCCAAAATC TCAAGGGCAA TGGCATAACA ATAGTTACTA
 2451 TGGTCTTATT TCTGCAGAAC ATCCCTTCCTT AAACCTGGGTG CTTCTTACAA
 2501 GACCTCTGGC TCAAGCTGG GATCTTTCAG GTTTTATTTC CGCAGAAATTC
 2551 CTAGGTGGTT GGCAAAGTAA GTTCACAGAA ACTGGAGATC TGCAACGTAG
 2601 CTTTAGTAGA GGTAAAGGGT ACAATGTTTC CCTACCGATA GGATGTTCTT
 2651 CTCATGGTT CACACCATT AAGAAGGCTC CTTCTACACT GACCACCAA
 2701 CTTGCCTACA AGCCTGATAT CTATCGTGTCA AACCCTCACA ATATTGTGAC
 2751 TGTGCTCTCA AACCAAGAGA GCACCTCGAT CTCAGGAGCA AATCTACGCC
 2801 GCCACGGTTT GTTGTACAA ATCCATGATG TAGTAGATCT CACCGAGGAC
 2851 ACTCAGGCCT TTCTAAACTA TACCTTGAC GGGAAAAATG GATTACAAA
 2901 CCACCGAGTG TCTACAGGAC TAAAATCCAC ATTTTAA

The PSORT algorithm predicts an outer membrane location (0.940).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 18A. The recombinant protein was used to immunise mice, whose sera were used in an immunoblot analysis blot (Figure 18B) and for FACS analysis (Figure 18C). A his-tagged protein was also expressed.

The cp6737 protein was also identified in the 2D-PAGE experiment (Cpn0454) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6737 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

35 Example 19

The following *C.pneumoniae* protein (PID 4377090) was expressed <SEQ ID 37; cp7090>:

1 **MNIHSLWKLC TLLALLALPA** CSLSPNYGWE DSCNTCHHTR RKKPSSFGFV
 51 PLYTEEDFNP NFTFGEYDSK EEKQYKSSQV AAFRNITFAT DSYTIKGEEN
 101 LAILTNLVHY MKKNPKATLY IEGHTDERGA ASYNLALGAR RANAIKEHLR
 151 KQGISADRLS TISYGKEHPL NSGHNELAWQ QRNRTEFKIH AR*

A predicted signal peptide is highlighted.

The cp7090 nucleotide sequence <SEQ ID 38> is:

1 ATGAATATAC ATTCCCTATG GAAACTTTGT ACTTTATTGG CTTTACTTG
 51 ATTGCCAGCA TGTAGCCTTT CCCCTAATTA TGGCTGGAG GATTCCTGTA
 101 ATACATGCCA TCATACAAGA CGAAAAAAGC CTTCTTCTTT TGGCTTTGTT
 151 CCTCTCTATA CCGAAGAGGA CTTTAACCCCT AATTTTACCT TCGGTGAGTA
 201 TGATTCCAAA GAAGAAAAAC AATACAAGTC AAGCCAAGTT GCAGCATTT
 251 GTAATATCAC CTTTGCTACA GACAGCTATA CAATTAAAGG TGAAGAGAAC
 301 CTTGCGATTG TCACGAACCTT GGTCACTAC ATGAAGAAAA ACCCGAAAGC
 351 TACACTGTAC ATTGAAGGGC ATACTGACGA GCGTGGAGCT GCATCCTATA
 401 ACCTTGCTTT AGGAGCACGA CGAGCCAATG CGATTAAGA GCATCTCCGA
 451 AAGCAGGGAA TCTCTGCAGA TCGTCTATCT ACTATTCCCT ACGGAAAAGA

501 ACATCCTTTA AATTGGGAC ACAACGAAC AGCATGGCAA CAAAATGCC
 551 GTACAGAGTT TAAGATTCA GCACGCTAA

The PSORT algorithm predicts an outer membrane location (0.790).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 19A.

- 5 A his-tagged protein was also expressed. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 19B) and for FACS analysis.

These experiments show that cp7090 is useful immunogen. These properties are not evident from the sequence alone.

Example 20

- 10 The following *C.pneumoniae* protein (PID 4377091) was expressed <SEQ ID 39; cp7091>:

1 **MLRQLCFQVF FFCFASLVYA** EELEVVRSE HITLPIEVSC QTDTKDPKIQ
 51 KYLSSLTEIF CKDIALGDCL QPTAASKESS SPLAISLRlh VPQLSVVLLQ
 101 SSKTPQTLCs FTISQNLSD RQKIHHADT VHYALTGIPG ISAGKIVFAL
 151 SSLGKDQKLK QGELWTTDYD GKNLAPLTTE CSLSITPKWV GVGSNFPYLY
 201 VSYKYGVPKI FLGSLNTEG KKVLPLKGNO LMPTFSPRKK LLAFVADTYG
 251 NPDLFIQPFS LTSGPMGRPR RLLNENFGTQ GNPSFNPEGS QLVFISNKDG
 301 RPRLYIMSLD PEPOQAPRLLT KKYNRSSCPA WSPDGKKIAF CSVIKGVRQI
 351 CIYDLSSGED YQLTTSPTNK ESPSWAIDSR HLVFSAGNAE ESELYLISLV
 401 TKKTNKIAIG VGEKRFPSWG AFPQQPIKRT L*

- 20 A predicted signal peptide is highlighted.

The cp7091 nucleotide sequence <SEQ ID 40> is:

1 ATGTTACGGC AACTATGCTT CCAAGTTTT TTCTTTTGCT TCGCATCGCT
 51 AGTCTATGCT GAAGAATTAG AAGTGTGTTGT CCGTTCCGAA CATATCACGC
 101 TCCCTATTGA GGTCTCTTCG CAGACCGATA CGAAAGATCC AAAAATACAG
 151 AAATACCTCA GCTCGCTAAC GGAGATATTT TGCAAGGACA TTGCCCTTAGG
 201 AGATTGTCTA CAACCCACAG CGGCTTCTAA AGAATCGCTA TCTCCTTTAG
 251 CAATATCTTT ACGGTTGCAT GTACCTCAGC TATCTGTAGT GCTTTTACAG
 301 TCTTCAAAAA CTCCTCAAC CTTATGTTCT TTTACTATTT CTCAAATCT
 351 TTCTGTAGAT CGTCAAAAAA TCCATCACGC TGCTGATACA GTTCATTACG
 401 CCCTCACAGG GATTCCTGGA ATCAGTGCTG GGAAAATTGT TTTTGTCTA
 451 AGTTCTTTAG GAAAAGATCA AAAGCTCAAG CAAGGAGAAT TATGGACTAC
 501 AGATTACGAT GGGAAAACC TCGCCCCCTT AACCAACAGAA TGTTCGCTCT
 551 CTATAACTCC AAAATGGGT GGTGTGGGAT CAAATTTCCT CTATCTCTAT
 601 GTTTCTGTATA AGTATGGTGT GCCTAAAAATT TTTCTTGGTT CCCTAGAGAA
 651 CACTGAAGGT AAAAAAGTCC TTCCGTTAAA AGGCAACCAA CTCATGCCA
 701 CGTTTCTC AAGAAAAAAAG CTTTAGCTT TCGTTGCTGA TACGTATGGA
 751 AATCTGATT TATTTATTCA ACCGTTCTCA CTAACCTCAG GACCTATGGG
 801 TCGCCCACGT CGCCTCCCTA ATGAGAATTG CGGGACTCAA GGGAAATCCCT
 851 CCTTCAACCC TGAAGGATCC CAGCTTGTCT TTATATCGAA CAAAGACGGC
 901 CGTCCCGCGTC TTTATATTAT GTCCCTCGAT CCTGAACCCC AAGCACCTCG
 951 CTTGCTGACA AAAAAATACA GAAATAGCAG TTGCCCCGCA TGGTCTCCAG
 1001 ATGGTAAAAA AATAGCCTTC TGCTCTGTAA TTAAAGGGT GCGACAAATT
 1051 TGTATTTACG ATCTCTCCCT TGGAGAGGAT TACCAACTCA CTACGTCTCC
 1101 CACAAATAAA GAGAGTCCTT CTTGGGCTAT AGACAGCCGT CATCTGTCT
 1151 TTAGTGCAGGG GAATGCTGAA GAATCAGAGT TATATTTAAT CAGTCTAGTC
 1201 ACCAAAAAAA CTAACAAAT TGCTATAGGA GTAGGAGAAA AACGGTTCCC
 1251 CTCCCTGGGGT GCTTCCCTC AGCAACCGAT AAAGAGAACAA CTATGA

The PSORT algorithm predicts an inner membrane location (0.109).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 20A.

- 50 A his-tagged protein was also expressed. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 20B) and for FACS analysis.

These experiments show that cp7091 is a useful immunogen. These properties are not evident from the sequence alone.

Example 21

The following *C.pneumoniae* protein (PID 4376260) was expressed <SEQ ID 41; cp6260>:

```

5      1  MRFSLCGFPL VFSFTLLSVF DTSILSATTIS LTPEDSFHGD SQNAERSYNV
      51  QAGDVYSLTG DVSI SVNVDNS ALNKACFNVT SGSVTFAGNH HGLYFNNI
10     101 GTTKEGAVLC CQDPQATARF SGFSTLSFIQ SPGDIKEQGC LYSKNALML
15     151 NNYVVRFEQN QS GTKGGAIIS GANVTIVGNY DSVSFYQNAAA TFGGAIHSSG
20     201 PLQIAVNQAE IRFAQNTAKN GSGGALYSDG DIDIDQNAYV LFRENEALTT
25     251 AIGKGGAVCC LPTSGSSTPV PIVTFSDNKQ LVFERNH SIM GGGAIYARKL
30     301 SISSGGPTLF INNISYANSQ NLGGAI AIDT GGEISLSAEK GTITFQGNRT
35     351 SLPFLNGIHL LQNAKFLKLQ ARNGYSIEFY DPITSEADGS TQLNINGDPK
40     401 NKEYTGTILF SGEKS LANDP RDKFKSTIPQN VNLSAGYLVI KEGAETVSK
45     451 FTQSPGSHLV LDLGTKL IAS KEDIAITGLA IDIDSLSSSS TAAVIKANTA
50     501 NKQISVTDI ELISPTGNAY EDLRRMRNSQT FPILLSLEPGA GGSVTVTAGD
55     551 FLPVSPHYGF QGNWKLA WIG TGNKVGEFFW DKINYKPRPE KEGNLVPNIL
60     601 WGNADVRSL MQVQETHASS LQTDRLGLWID GIGNFFFHVA SEDNIRYRHN
65     651 SGGYVLSVNN EITPKHYTSM AFSQLFSRDK DYAVSNNEYR MYLGSYLYQY
70     701 TTSLGNIFRY ASRNPNVNVG ILSRRFLQNP LMIFHFLCAY GHATNDMKT
75     751 YANFPVMKNS WRNNCWAIEC GGSMPLLVE NGRLFQGAIP FMKLQLVYAY
80     801 QGDFKETTAD GRRFSNGSLT SISVPLGIRF EKLALSQDVL YDFSF SYIPD
85     851 IFRKDPSCA ALVISGDSWL VPAAHVSRHA FVGSGTGRYH FNDYTELLCR
90     901 GSIECRPHAR NYNINC GSKF RF*

```

A predicted signal peptide is highlighted.

25 The cp6260 nucleotide sequence <SEQ ID 42> is:

```

1  ATGC GATT TT CGCT CTGCGG ATT CCTCTCTA GTT TTTT CTT TTACATTGCT
5  51  CTCAGTCTTC GACACTT CTT TGAGT GCTAC TACGATT TCT TTAACCCAG
10  101 AAGATAGTT TCATGGAGAT AGTCAGAATG CAGAACGTT TTATAATGTT
15  151 CAAGCTGGGG ATGTCTATAG CCTTACTGGT GATGTC TCAA TATCTAACGT
20  201 CGATAACTCT GCATTAATAA AAGCCTGCTT CAATGTGACC TCAGGAAGTG
25  251 TGACGTT CGC AGGAAATCAT CATGGTTAT ATTTAATAA TATTCCTCA
30  301 GGAAC TACAA AGGAAGGGC TGTACTTTGT TGCCAAGATC CTCAAGAAC
35  351 GGCACGTT TCTGGTTCT CCACGCTCTC TTTTATTCAAG AGCCCCGGAG
40  401 ATATTAAAGA ACAGGGATGT CTCTATTCAA AAAATGCACT TATGCTCTTA
45  451 AACAA TTATG TAGT GCGTTT TGAACAAA CAAAGTAAGA CTAAAGGCGG
50  501 AGCTATTAGT GGGCGGAATG TTACTATAGT AGGCAACTAC GATTCCGTCT
55  551 CTTTCTATCA GAATGCAGCC ACTTTTGGAG GTGCTATCCA TTCTTCAGGT
60  601 CCCCTACAGA TTGCAGTAAA TCAGGCAGAG ATAAGATTG CACAAAATAC
65  651 TGCCAAGAAT GGTTCTGGAG CGGCCTTGTA CTCCGATGGT GATATTGATA
70  701 TTGATCAGAA TGCTTATGTT CTATTCGAG AAAATGAGGC ATTGACTACT
75  751 GCTATAGGTA AGGGAGGGC TGTCTGTTGT CTTCCCAC TT CAGGAAGTAG
80  801 TACTCCAGTT CCTATTGTA CTTTCTCTGA CAATAAACAG TTAGTCTTTG
85  851 AAAGAAACCA TTCCATAATG GGTGGCGGAG CCATTATGC TAGGAAACTT
90  901 AGCATCTCTT CAGGAGGTCC TACTCTATT ATCAATAATA TATCATATGC
95  951 AAATTCGCAA AATT TAGGTG GAGCTATTGC CATTGATACT GGAGGGGAGA
100 1001 TCAGTTTATC AGCAGAGAAA GGAACAATTA CATTCCAAGG AAACCGGACG
105 1051 AGCTTACCGT TTTGGAATGG CATCCATCTT TTACAAAATG CTAAATTCCCT
110 1101 GAAATTACAG GCGAGAAATG GATACTCTAT AGAATT TAT GATCCTATT
115 1151 CTTCTGAAGC AGATGGTCT ACCCAATTGA ATATCAACGG AGATCCTAAA
120 1201 AATAAAAGAGT ACACAGGGAC CATACTCTT TCTGGAGAAA AGAGTCTAGC
125 1251 AAACGATCCT AGGGATT TTA AATCTACAT CCTCTCAGAAC GTCAACCTGT
130 1301 CTGCAGGATA CTTAGTTATT AAAGAGGGGG CGGAAGTCAC AGTTTCAAAA
135 1351 TTCACCGAGT CTCCAGGATC GCATTAGTT TTAGATTAG GAACCAAAC
140 1401 GATAGCCTCT AAGGAAGACA TTGCCATCAC AGGCCTCGCG ATAGATATAG
145 1451 ATAGCTTAAG CTCATCCTCA ACAGCAGCTG TTATTAAGC AAACACCGCA
150 1501 AATAAAACAGA TATCCGTGAC GGACTCTATA GAACCTATCT CGCCTACTGG
155 1551 CAATGCCAT TGAAGATCTCA GAATGAGAAA TTCACAGACG TTCCCTCTGC
160 1601 TCTCTTTAGA GCCTGGAGCC GGGGGTAGTG TGACTGTAAC TGCTGGAGAT
165 1651 TTCC TACCGG TAAGTCCCCA TTATGGTTT CAAGGCAATT GGAAATTAGC
170 1701 TTGGACAGGA ACTGGAAACA AAGTTGGAGA ATTCTCTGG GATAAAATAA

```

1751 ATTATAAGCC TAGACCTGAA AAAGAAGGAA ATTTAGTTCC TAATATCTTG
 1801 TGGGGGAATG CTGTAGATGT CAGATCCTTA ATGCAGGTTC AAGAGACCCA
 1851 TGCATCGAGC TTACAGACAG ATCGAGGGCT GTGGATCGAT GGAATTGGGA
 1901 ATTTCTTCCA TGTATCTGCC TCCGAAGACA ATATAAGGT ACGTCATAAC
 1951 AGCGGTGGAT ATGTTCTATC TGAAATAAT GAGATCACAC CTAAGCACTA
 2001 TACTTCGATG GCATTTCCC AACCTTTAG TAGAGACAAG GACTATGCGG
 2051 TTTCCAACAA CGAACATACAGA ATGTAATTAG GATCGTATCT CTATCAATAT
 2101 ACAACCTCCC TAGGGAAATAT TTTCCGTAT GCTTCGCGTA ACCCTAATGT
 2151 AAACGTCGGG ATTCTCTCAA GAAGGTTTCTC TCAAATCCT CTTATGATTT
 2201 TTCATTTTT GTGTGCTTAT GGTATGCCA CCAATGATAT GAAAACAGAC
 2251 TACCGCAAATT TCCCTATGGT GAAAAACAGC TGGAGAAACA ATTGTTGGGC
 2301 TATAGAGTGC GGAGGGAGCA TGCCCTCTATT GGTATTGAG AACGGAAGAC
 2351 TTTCCAAGG TGCCATCCC TTTATGAAAC TACAATTAGT TTATGCTTAT
 2401 CAGGGAGATT TCAAAGAGAC GACTGCAGAT GGCGTAGAT TTAGTAATGG
 2451 GAGTTTAACA TCGATTCTG TACCTCTAGG CATAKGCTT GAGAACGCTGG
 2501 CACTTTCTCA GGATGTACTC TATGACTTTA GTTTCTCTA TATTCTGTAT
 2551 ATTTTCCGTA AGGATCCCTC ATGTGAAGCT GCTCTGGTGA TTAGCGGAGA
 2601 CTCCTGGCTT GTTCCGGCAG CACACGTATC AAGACATGCT TTTGTAGGGA
 2651 GTGGAACGGG TCGGTATCAC TTTAACGACT ATACTGAGCT CTTATGTCGA
 2701 GGAAGTATAAG AATGCCGCC CCATGCTAGG AATTATAATA TAAACTGTGG
 2751 AAGCAAATT CGTTTTAG

The PSORT algorithm predicts an outer membrane location (0.921).

The protein was expressed in *E.coli* and purified both as a his-tag and GST-fusion product. The GST-fusion is shown in Figure 21A. This recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 21B) and for FACS analysis (Figure 21C).

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6260 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

30 Example 22

The following *C.pneumoniae* protein (PID 4376456) was expressed <SEQ ID 43; cp6456>:

1 MSSPVNNTPS APNIPIPAPT TPGIPTTKPR SSFIEKVIIIV AKYILFAIAA
 51 TSGALGTILG LSGALTPGIG IALLVIFVFVS MVLLGLILKD SISGGEERRL
 101 REEVSRFTSE NQRLTIVITTT LETEVKDLKA AKDQLTLEIE AFRNENGNLK
 151 TTAEDLEEQQV SKLSEQEAL ERINQLIQAN AGDAQEISSE LKKLISGWDS
 201 KVVEQINTSI QALKVLLGQE WVQEAQTHVK AMQEIQALQ AEILGMHQS
 251 TALQKSVENL LVQDQALTRV VGGLESENK LSQACSLRQ EIEKLAQHET
 301 SLQQRIDAML AQEQNLAEQV TALEKMKQEA QKAESEFIAC VRDRTFGRRE
 351 TPPPTTPVVE GDESQEEDEG GTPPPVSPQSS PVDRATGDQ *

40 The cp6456 nucleotide sequence <SEQ ID 44> is:

1 ATGTCATCTC CTGAAATAA CACACCCTCA GCACCAAACA TTCCAATACC
 51 AGCGCCCACG ACTCCAGGT A TCCCTACAAC AAAACCTCGT TCTAGTTCA
 101 TTGAAAAGGT TATCATGTG GCTAAGTACA TACTATTGTC AATTGCGAGCC
 151 ACATCAGGAG CACTCGGAAC AATTCTAGGT CTATCTGGAG CGCTAACCCC
 201 AGGAATAGGT ATTGCCCTTC TTGTTATCTT TTTGTTTCT ATGGTGCTTT
 251 TAGGTTTAAT CCTTAAAGAT TCTATAAGTG GAGGAGAAGA ACGCAGGCTC
 301 AGAGAAGAGG TCTCTCGATT TACAAGTGAG AATCAACGGT TGACAGTCAT
 351 AACCCACAACA CTTGAGACTG AAGTAAAGGA TTTAAAGCA GCTAAAGATC
 401 AACTTACACT TGAAATCGAA GCATTTAGAA ATGAAACGG TAATTAAAAA
 451 ACAACTGCTG AGGACTTAGA AGAGCAGGTT TCTAAACTTA GCGAACAAATT
 501 AGAAGCACTA GAGCGAATTA ATCAACTTAT CCAAGCAAAC GCTGGAGATG
 551 CTCAAGAAAT TTGCTCTGAA CTAAGAAAT TAATAAGCGG TTGGGATTCC
 601 AAAGTTGTTG AACAGATAAA TACTTCTATT CAAGCATGTA AAGTGTATT
 651 GGGTCAAGAG TGGGTGCAAG AGGCTCAAAC ACACGTTAAA GCAATGCAAG
 701 AGCAAATTCA AGCATTGCAA GCTGAAATTTC TAGGAATGCA CAATCAATCT

5 751 ACAGCATTGC AAAAGTCAGT TGAGAATCTA TTAGTACAAG ATCAAGCTCT
 801 AACAAAGAGTA CTAGGTGAGT TGTAGAGTC TGAGAACAAAG CTAAGCCAAG
 851 CTITGTTCTGC GCTACGTCAA GAAATAGAAA AGTTGGCCCA ACATGAAACA
 901 TCTTTGCAAC AACGTATTGA TGCATGCTA GCCCAAGAGC AAAATTGGC
 951 AGAGCAGGTC ACAGGCCCTG AAAAAATGAA ACAAGAACGCT CAGAAGGCTG
 1001 AGTCCGAGTT CATTGCTTGT GTACGTGATC GAACCTTCGG ACGTCGTGAA
 1051 ACACCTCAC CAACAACACC TGAGTTGAA GGATGATGAAA GTCAAGAAGA
 1101 AGACGAAGGA GGTACTCCCC CAGTATCACA ACCATCTTCA CCCGTAGATA
 1151 GAGCAACAGG AGATGGTCAG TAA

10 The PSORT algorithm predicts inner membrane (0.127).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 22A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 22B) and for FACS analysis (Figure 22C). A his-tag protein was also expressed.

15 These experiments show that cp6456 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 23

The following *C.pneumoniae* protein (PID 4376729) was expressed <SEQ ID 45; cp6729>:

20 1 MKIPLHKLLL SSTLVTPILL SIATYGDAS LSPTDSFDGA GGSTFTPKST
 51 ADANGTNYVL SGNVYINDAG KGTALTGCCF TETTGDLTFT GKGYSFSFNT
 101 VDAGSNAGAA ASTTADKALT FTGFSNLSFI AAPGTTVASG KSTLSSAGAL
 151 NLTDNGTILF SQNVSNEANN NGGAITTAKTL SISGNTSSIT FTSNSAKKLG
 201 GAIYSSAAAS ISGNTGQLVF MNNKGETGGG ALGFEASSSI TQNSSLFFSG
 251 NTATDAAGKG GAIYCEKTGE TPTLTISGNK SLTFAENSSV TQGGAICAHG
 301 LDLSAAGPTL FSNNRCCNTA AGKGGAAIA DSGSLSLSAN QGDITFLGNT
 351 LTSTSAPTST RNAIYLGSAA KITNLRAAQG QSIYFYDPIA SNTTCASDVL
 401 TINQPDSNSP LDYSGTIVFS GEKLSADEAK AADNFTSILK QPLALASGTL
 451 ALKGNVELDV NGFTQTEGST LLMQPGTKLK ADTEAISLTK LVVDLSALEG
 501 NKSVISIETAG ANKTITLTSP LVFQDSSGNF YESHTINQAF TQPLUVFTAA
 551 TAASDIYIDA LLTSPVQTPE PHYGYQGHWE ATWADTSTAK SGTMTWVTG
 601 YNPNPERRAS VVPDSLWASF TDIRTLQQIM TSQANSIYQQ RGLWASGTAN
 651 FFHKDKSGTN QAFRHKSYGY IVGGSAEDFS ENIFSVAFQC LFGKDKDLF
 701 VENTSHNYLA SLYLQHRAFL GGLPMPSFGS ITDMLDIPL ILNAQLSYSY
 751 TKNDMDTRYT SYPEAQGSWT NNSGALELGG SLALYLPKEA PFFQGYFPFL
 801 KFQAVYSRQQ NFKESGAEAR AFDDGDLVNC SIPVGIRLEK ISEDEKNNFE
 851 ISLAYIGDVTY RKNPRSRSTSL MVSGASWTSK CKNLARQAFL ASAGSHLTL
 901 PHVELSGEAA YELRGSAHIY NVDCGLRYSF *

A predicted signal peptide is highlighted.

The cp6729 nucleotide sequence <SEQ ID 46> is:

40 1 ATGAAAATAC CCTTGACCAA ACTCCTGATC TCTTCGACTC TTGTCACTCC
 51 CATTCTATTG AGCATTGCAA CTTACGGAGC AGATGCTTCT TTATCCCCTA
 101 CAGATAGCTT TGATGGAGCG GGCGGCTCTA CATTACTCC AAAATCTACA
 151 GCAGATGCCA ATGGAACGAA CTATGTCTTA TCAGGAAATG TCTATATAAA
 201 CGATGCTGGG AAAGGCACAG CTTAACAGG CTGCTGCTTT ACAGAAACTA
 251 CGGGTGATCT GACATTACT GGAAAGGGAT ACTCATTTC ATTCAACACG
 301 GTAGATGCGG GTTCGAATGC AGGAGCTGCG GCAAGCACAA CTGCTGATAA
 351 AGCCCTAACAA TTCACAGGAT TTTCTAACCT TCCCTTCATT GCAGCTCCTG
 401 GAACTACAGT TGCTTCAGGA AAAAGTACTT TAAGTTCTGC AGGAGCCTTA
 451 AATCTTACCG ATAATGGAAC GATTCCTTT AGCCAAAACG TCTCCAATGA
 501 AGCTAAATAAC AATGGGGAG CGATCACCAC AAAAATCTTT TCTATTCTG
 551 GGAATACCTC TTCTATAAACC TTCACTAGTA ATAGCGCAA AAAATTAGGT
 601 GGAGCGATCT ATAGCTCTGC GGCTGCAAGT ATTCAGGAA ACACCGCCA
 651 GTTAGTCTTT ATGAATAATA AAGGAGAAAC TGGGGGTGGG GCTCTGGGCT
 701 TTGAAGCCAG CTCCTCGATT ACTCAAAATA GCTCCCTTT CTTCTCTGGA
 751 AACACTGCAA CAGATGCTGC AGGCAAGGGC GGGGCCATT ATTGTGAAAA
 801 AACAGGAGAG ACTCCTACTC TTACTATCTC TGGAAATAAA AGTCTGACCT
 851 TCGCCGAGAA CTCTTCAGTA ACTCAAGGCG GAGCAATCTG TGCCCATGGT

	901	CTAGATCTT CCGCTGCTGG CCCTACCCCTA TTTTCAAATA ATAGATGC GG
	951	GAACACAGCT GCAGGCAAGG CGGGCGCTAT TGCAATTGCC GACTCTGGAT
5	1001	CTTAAAGTCT CTCTGCAAAT CAAGGAGACA TCACGTTCT TGGCAACACT
	1051	CTAACCTCAA CCTCCGCGCC AACATCGACA CGGAATGCTA TCTACCTGGG
	1101	ATCGTCAGCA AAAATTACGA ACTTAAGGGC AGCCCCAAGGC CAATCTATCT
	1151	ATTCTATGA TCCGATTGCA TCTAACACCA CAGGAGCTC AGACGTTCTG
	1201	ACCATCAACC AACCGGATAG CAACTCGCT TTAGATTATT CAGGAACGAT
	1251	TGTATTTCT GGGGAAAAGC TCTCTGAGA TGAGCGAAA GCTGCTGATA
10	1301	ACTTCACATC TATTTAAAG CAACCATTTG CTCTAGCCTC TGGAACCTTA
	1351	GCACCTCAAAG GAAATGTCGA GTTAGATGTC AATGGTTCA CACAGACTGA
	1401	AGGCTCTACA CTCCCTCATGC ACCAGGAAC AAAGCTCAA GCAGATACTG
	1451	AAGCTATCAG TCTTACCAAA CTTGTCGTTG ATCTTCTGC CTTAGAGGGA
	1501	AATAAGAGTG TGTCCATTGA AACAGCAGGA GCCAACAAAA CTATAACTCT
15	1551	AACCTCTCCT CTTGTTTCC AAGATAGTAG CGGCAATT TT TATGAAAGCC
	1601	ATACGATAAA CCAAGCCTTC ACGCAGCCTT TGGTGGTATT CACTGCTGCT
	1651	ACTGCTGCTA GCGATATT TA TATCGATGCG CTTCTCACCT CTCCAGTACA
	1701	AACTCCAGAA CCTCATTACG GGTATCAGGG ACATTGGGAA GCCACCTGGG
	1751	CAGACACATC AACTGCAAA TCAGGAAC TA TGACTTGGGT AACTACGGGC
20	1801	TACAACCCCTA ATCCGTAGCG TAGAGCTTCC GTAGTTCCCG ATTCAATTATG
	1851	GGCATCCTTT ACTGACATTC GCACTCTACA GCAGATCATG ACATCTCAAG
	1901	CGAATAGTAT CTATCAGCAA CGAGGACTCT GGGCATCAGG AACTGCGAAT
	1951	TTCTTCCATA AGGATAAAATC AGGAACTAAC CAAGCATTCC GACATAAAAG
	2001	CTACGGCTAT ATTGTTGGAG GAAGTGTGCTGA AGATTTTCT GAAAATATCT
25	2051	TCAGTGTAGC TTTCTGCCAG CTCTCGGTA AAGATAAAAGA CCTGTTTATA
	2101	GTGAAAATA CCTCTCATAA CTATTAGCG TCGCTATACC TGCAACATCG
	2151	AGCATTCCCTA GGAGGACTTC CCATGCCCTC ATTGGAAGT ATCACCGACA
	2201	TGCTGAAAGA TATTCCCTC ATTGGAATG CCCAGCTAAG CTACAGCTAC
	2251	ACTAAAAATG ATATGGATAC TCGCTATACT TCCTATCCTG AAGCTCAAGG
30	2301	CTCTTGGGACC AATAACTCTG GGGCTCTAGA GCTCGGAGGA TCTCTGGCTC
	2351	TATATCTCCC TAAAGAAGCA CCGTTCTTCC AGGGATATT CCCCTCTTA
	2401	AAGTTCCAGG CAGTCTACAG CGCCAACAA AACTTTAAAG AGAGTGGCGC
	2451	TGAAGCCCGT GCTTTGATG ATGGAGACCT AGTGAAGTGC TCTATCCTG
	2501	TCGGCATTCTG GTTAGAAAAA ATCTCCGAAG ATGAAAAAAA TAATTTCGAG
35	2551	ATTTCTCTAG CCTACATTGG TGATGTTGAT CGTAAAAATC CCCGTTCGCG
	2601	TACTTCTCTA ATGGTCAGTG GACCTCTTG GACTTCGCTA TGTAACACC
	2651	TCGCACGACA AGCCTCTTA GCAAGTGTG GAAGCCATCT GACTCTCTCC
	2701	CCTCATGTAG AACTCTCTGG GGAAGCTGCT TATGAGCTC GTGGCTCAGC
	2751	ACACATCTAC AATGTAGATT GTGGCTAAG ATACTCATTC TAG

The PSORT algorithm predicts outer membrane (0.927).

- 40 The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 23A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 23B) and for FACS analysis (Figure 23C). A his-tag protein was also expressed.

The cp6729 protein was also identified in the 2D-PAGE experiment (Cpn0446) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

- 45 These experiments show that cp6729 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 24

The following *C.pneumoniae* protein (PID 4376849) was expressed <SEQ ID 47; cp6849>:

50	1	MSKLIRRVT VLALTSMASC FASGGIEAAV AESLITKIVA SAETKPAPVP
	51	MTAKKVRLLVR RNKQPVEQKS RGAFCDKEFY PCEEGRCQPV EAQQESCYGR
	101	LYSVKVNDDC NVEICQSVPY YATVGSPYPI EILAIGKKDC VDVVITQQLP
	151	CEAEFVSSDP ETTPTSDGKL VKIDRLGAG DKCKITVVVK PLKEGCCFTA
	201	ATVCACPELR SYTKCGQPAI CIKQEGPDCA CLRPCVCYKI EVVNIGSAIA
	251	RNVTVDNPVP DGYSHASGQR VLSFNLDMDR PGDKKVFTVE FCPQRRQQT
55	301	NVATVTVYCGG HKCSANVTTV VNEPCVQVNI SGADWSYVCK PVEYSISVSN
	351	PGDLVLHDVV IQDTLPMSGVT VLEAPGGEIC CNKVVWRIKE MCPGETLQFK

-66-

401 LVVKAQVPGR FTNQVAVTSE SNCGTCTSCA ETTTHWKGLA ATHMCVLDTN
 451 DPICVGENTV YRICVTNRGS AEDTNVSLIL KFSKELQPIA SSGPTKGTIS
 501 GNTVVFDALP KLGSKESVEF SVTLKGIAPG DARGEAILSS DTLTSPVSDT
 551 ENTHVY*

- 5 A predicted signal peptide is highlighted.

The cp6849 nucleotide sequence <SEQ ID 48> is:

1 ATGTCCAAAC TCATCAGACG AGTAGTTACG GTCCCTGGCG TAACGAGTAT
 51 GCGCAGTTGC TTTGCCAGCG GGGGTATAGA GGCGCTGTA GCAGAGTC
 101 TGATTACTAA GATCGTCGCT AGTGCAGAAA CAAAGCCAGC ACCTGTTCT
 151 ATGACAGCGA AGAAGGTTAG ACTTGTCCGT AGAAATAAAC AACCAAGTTGA
 201 ACAAAAAAGC CGTGGTGCTT TTTGTGATAA AGAATTTTAT CCCTGTGAAG
 251 AGGGACGATG TCAACCTGTA GAGGCTCAGC AAGAGTCTTG CTACGGAAGA
 301 TTGTTATTCTG TAAAAGTAAA CGATGATTGC AACGTAGAAA TTTGCCAGTC
 351 CGTTCCAGAA TAGCTACTG TAGGATCTCC TTACCCCTATT GAAATCCTTG
 401 CTATAGGCAA AAAAGATTGT GTTGATGTTG TGATTACACA ACAGCTACCT
 451 TCGAAGCTG AATTCTGAAG CAGTGATCCA GAAACAACTC CTACAAGTGA
 501 TGGGAAATTG GTCTGGAAAA TCGATCGCCT GGGTGCAGGA GATAAAATGCA
 551 AAATTACTGT ATGGGTAAAA CCTCTTAAAG AAGGTTGCTG CTTCACAGCT
 601 GCTACTGTAT GTGCTTGCCC AGAGCTCCGT TCTTATACTA AATGCGGTCA
 651 ACCAGCCATT TGTATTAAGC AAGAAGGACC TGACTGTGCT TGCCTAAAGAT
 701 GCCCTGTATG CTACAAAATC GAAGTAGTGA ACACAGGATC TGCTATTGCC
 751 CGTAACGTAACCTGTAGATAA TCCCTGTTCCC GATGGCTATT CTCATGCATC
 801 TGGTCAAAGA GTTCTCTCTT TTAACTTAGG AGACATGAGA CCTGGCGATA
 851 AAAAGGTATT TACAGTTGAG TTCTGCCCTC AAAGAAGAGG TCAAATCACT
 901 AACGTTGCTA CTGTAACCTA CTGCGGTGGA CACAAATGTT CTGCAAATGT
 951 AACTACAGTT GTTAATGAGC CTTGTGTACA AGTAAATATC TCTGGTGCTG
 1001 ATTGGTCTTA CGTATGTAAA CCTGTGGAGT ACTCTATCTC AGTATCGAAT
 1051 CCTGGAGACT TGGTTCTTCA TGATGTGCTG ATCCAAGATA CACTCCCTTC
 1101 TGGTGTACAGTCA GTACTCGAAG CTCCCTGGTGG AGAGATCTGC TGTAAATAAG
 1151 TTGTTTGGCG TATTAAAGAA ATGTGCCAG GAGAAACCT CCAGTTTAAA
 1201 CTTGTAGTGA AAGCTCAAGT TCCCTGGAAGA TTCACAAATC AAGTTGCAGT
 1251 AACTAGTGAG TCTAACTGCG GAACATGTAC ATCTTGCGCA GAAACAAACAA
 1301 CACATTGGAA AGGTCTTGCA GCTACCCATA TGTGCGTATT AGACACAAAT
 1351 GATCCTATCT GTGTAGGAGA AAATACTGTC TATCGTATCT GTGTAACCAA
 1401 CCGTGGTTCTG GCTGAAGATA CTAACGTATC TTTAATCTTG AAGTTCTCAA
 1451 AAGAAACTTCA GCCAATAGCT TCTTCAGGTC CAAACTAAAGG AACGATTTC
 1501 GGTAAATACCG TTGTTTTCGA CGCTTTACCT AAACCTGGTT CTAAGGAATC
 1551 TGAGAGTTT TCTGTACCT TGAAAGGTAT TGCTCCCGGA GATGCTCGCG
 1601 GCGAAGCTAT TCTTTCTTCT GATACACTGA CTTCACCAAGT ATCAGACACA
 1651 GAAAATACCC ACGTGTATTA A

The PSORT algorithm predicts periplasmic space (0.93).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 24A, and also as a his-tag protein. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 24B) and for FACS analysis (Figure 24C).

- 45 The cp6849 protein was also identified in the 2D-PAGE experiment (Cpn0557).

These experiments show that cp6849 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 25

The following *C.pneumoniae* protein (PID 4376273) was expressed <SEQ ID 49; cp6273>:

50 1 MGLFHLTFLFG LLLCSLPISL VAKFPESVGH KILYISTQST QQALATYLEA
 51 LDAYGDHDFF VLRKIGEDYL KQSIHSSDPQ TRKSTIIGAG LAGSSEALDV
 101 LSQAMETADP LQQLLVLSAV SGHLGKTSDD LLFKALASPY PViRLEAAYR
 151 LANLKNKTVI DHLHSFIHKL PEEIQCLSAA IFLRLETEES DAYIRDLLAA
 201 KKSAAIRSATA LQIGEYQQKR FLPTLRLNLLT SASPQDQEAI LYALGKLKD

-67-

5 251 QSYVNIKKQL QKPDVDVTLA AAOALIALGK EEDALPVKK QALEERPRAL
 301 YALRHLPSERI GIPIALPIFL KTKNSEAKLN VALALLELGC DTPKLLEYIT
 351 ERLVQPHYNE TLALSFSKGR TLQNWKRVNI IVPQDPQERE RLLSSTTRGLE
 401 EQILTFLFRL PKEAYLPCY KLLASQKTQL ATTAISFLSH TSHQEALDLL
 451 FQAAKLPGEF IIRAYADLAI YNLTKDPEKK RSLHDYAKKL IQETLLFVDT
 501 ENQRPHPSMP YLRYQVTPES RTKLMLDILE TLATSKSS ED IRLLIQLMTE
 551 GDAKNFPVLA GLLIKIVE*

A predicted signal peptide is highlighted.

The cp6273 nucleotide sequence <SEQ ID 50> is:

10 1 ATGGGACTAT TCCATCTAAC TCTCTTGGA CTTTTATTGT GTAGTCCTCC
 51 CATTTCCTT GTTGCTAAAT TCCCTGAGTC TGTAGGTCAT AAGATCCTTT
 101 ATATAAGTAC GCAATCTACA CAGCAGGCC TAGCAACATA TCTGGAAGCT
 151 CTAGATGCCT ACGGTGATCA TGACTTCTTC GTTTTAAGAA AAATCGGAGA
 201 AGACTATCTC AAGCAAAGCA TCCACTCCTC AGATCCGAA ACTAGAAAAAA
 251 GCACCATCAT TGGAGCAGGC CTGGCGGGAT CTTCAGAACG CTTGGACGTG
 301 CTCTCCCAAG CTATGAAAC TGCAAGACCCC CTGCAGCAGC TACTGGTTT
 351 ATCGGCAGTC TCAGGACATC TTGGGAAAC TTCTGAGCAC TTACTGTTA
 401 AAGCTTTAGC ATCTCCCTAT CCTGTCATCC GCTTAGAACG CGCCTATAGA
 451 CTTGCTAATT TGAAGAACAC TAAAGTCATT GATCATCTAC ATTCTTTCAT
 501 TCATAAGCTT CCCGAAGAAA TCCAATGCCT ATCTGCGCA ATATTCCTAC
 551 GCTTGGAGAC TGAAGAATCT GATGCTTATA TTCGGGATCT CTTAGCTGCC
 601 AAGAAAAGCG CGATTCCGGAG TGCCACAGCT TTGCAGATCG GAGAATACCA
 651 ACAAAAACGC TTTCTCCGA CACTTAGGAA TTGCTAAGC AGTGCCTCTC
 701 CTCAAGATCA AGAAGCTATT CTTTATGCTT TAGGGAAGCT TAAGGATGGT
 751 CAGAGCTACT ACAATATAAA AAAGCAATTG CAGAACGCTG ATGTTGGATGT
 801 CACTTTAGCA GCAGCTCAAG CTTTAATTGC TTGGGGAAA GAAGAGGACG
 851 CTCTTCCCGT GATAAAAAG CAAGCACTTG AGGAGCGGCC TCGAGCCCTG
 901 TATGCCCTAC GGCATCTACC CTCTGAGATA GGGATTCCGA TTGCCCCGCC
 951 GATATTCTCA AAAACTAAGA ACAGCGAACG CAAGTTGAAT GTAGCTTTAG
 1001 CTCTCTTCTAGA GTTAGGGTGT GACACCCCTA AACTACTGGA ATACATTACC
 1051 GAAAGGCTTG TCCAACCACA TTATAATGAG ACTCTAGGCT TGAGTTCTC
 1101 TAAGGGGCGT ACTTTACAAA ATTGGAAGCG GGTGAACATC ATAGTCCCTC
 1151 AAAGATCCCCA GGAGAGGGAA AGGTTGCTCT CCACAAACCG AGGTCTTGAA
 1201 GAGCAGATCC TTACGTTTCT CTTCCGCCCTA CCTAAAGAAG CTTACCTCCC
 1251 CTGTATTTAT AAGCTTTTGG CGAGTCAGAA AACTCAGCTT GCCACTACTG
 1301 CGATTTCTTT TTAAAGTCAC ACCTCACATC AGGAAGCCTT AGATCTACTT
 1351 TTCCAAAGCTG CGAAGCTTCC TGGAGAACCT ATCATCCGCG CCTATGCAGA
 1401 TCTTGCCTATT TATAATCTCA CCAAAGATCC TGAAAAAAA CGTTCTCTCC
 1451 ATGATTATGC AAAAAGCTA ATTCAAGGAAA CCTTGTATT TGTGGACACG
 1501 GAAAACCAAA GACCCCATCC CAGCATGCC TATCTACGTT ATCAGGTAC
 1551 CCCAGAAAGC CGTACGAAGC TCATGTTGGA TATTCTAGAG ACACATAGCCA
 1601 CCTCGAAGTC TTCCGAAGAT ATCCGTTTAT TGATACAACT GATGACGGAA
 1651 GGAGATGCAA AAAATTCTCC AGTCCTTGCA GGCTTACTCA TAAAAATTGT
 1701 GGAGTAA

45 The PSORT algorithm predicts a periplasmic location (0.922).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product, as shown in Figure 25A. The recombinant GST-fusion was used to immunise mice, whose sera were used in a Western blot (Figure 25B) and for FACS analysis (Figure 25C).

This protein also showed good cross-reactivity with human sera, including sera from patients with 50 pneumonitis.

These experiments show that cp6273 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 26

The following *C.pneumoniae* protein (PID 4376735) was expressed <SEQ ID 51; cp6735>:

5 1 MTILRNFLTC SALFLALPAA AQVVLHESD GYNGAINNKS LEPKITCYPE
 51 GTSYIFLDDV RISNVKHDQE DAGVFINRSG NLFFMGNRCN FTFHNLMTEG
 101 FGAAISNRVG DTTLTLNSNFs YLAFTSAPLL PQGQGAIYSL GSVMIEENSEE
 151 VTFCGNYSSW SGAAIYTPYL LGSKASRPSV NLSGNRYLVF RDNVSQGYGG
 201 AISTHNLTLT TRGPSCFENN HAYHDVNSNG GAIAIAPGGS ISISVKSGDL
 251 IFKGNTASQD GNTIHNSIHL QSGAQFKNLR AVSESGVYFY DPISHSESHK
 301 ITDLVINAPE GKETYEGTIS FSGLCLEDDHE VCAENLTSTI LQDVTLAGGT
 351 LSLSDGVTLQ LHSFKQEAASS TLTMSPGTTL LCSGDARVQN LHILIEDTDN
 401 FVPVRIRAED KDALVSLEKL KVAFEAYWSV YDFPQFKEAF TIPLLELLGP
 451 SFDSLLLGET TLERTQVITE NDAVRGFWSL SWEYPPSLD KDRRIPTKK
 501 TVFLTWNPEI TSTP*

A predicted signal peptide is highlighted.

The cp6735 nucleotide sequence <SEQ ID 52> is:

15 1 ATGACCATAAC TTGAAATTTC TCTTACCTGC TC GGCTTTAT TCCTCGCTCT
 51 CCCTGCAGCA GCACAAGTTG TATATCTTC TGAAAGTGAT GGTTATAACG
 101 GTGCTATCAA TAATAAAAGC TTAGAACCTA AAATTACCTG TTATCCAGAA
 151 GGAACCTTCTT ACATCTTTCT AGATGACGTG AGGATTTCGA ACGTTAACCA
 201 TGATCAAGAA GATGCTGGGG TTTTATAAA TCGATCTGGG AATCTTTTTT
 251 TCATGGGCAA CCGTTGCAAC TTCACTTTTC ACAACCTTAT GACCGAGGGT
 301 TTTGGCGCTG CCATTTGCAA CCGCGTTGGA GACACCACTC TCACTCTCTC
 351 TAATTTTCT TACTTAGCGT TCACCTCAGC ACCTCTACTA CCTCAAGGAC
 401 AAGGAGCGAT TTATAGTCTT GGTTCGGTGA TGATCGAAAA TAGTGAGGAA
 451 GTGACTTTCT GTGGGAACTA CTCTTCGTGG AGTGGAGCTG CGATTATAC
 501 TCCCTACCTT TTAGGTTCTA AGGCGAGTCG TCCCTCAGTA AATCTCAGCG
 551 GGAACCGCTA CCTGGTGTAG AGAGACAATG TGAGCCAAGG TTATGGCGGC
 601 GCCATATCTA CCCACAATCT CACACTCACG ACTCGAGGAC CTTCGTGTGTTT
 651 TGAAAATAAT CATGCTTATC ATGACGTGAA TAGTAATGGA GGAGCCATTG
 701 CCATTGCTCC TGGAGGATCG ATCTCTATAT CCGTGAAAAG CGGAGATCTC
 751 ATCTTCAAAG GAAATACAGC ATCACAAAGAC GGAAATACAA TACACAAC
 801 CATCCATCTG CAATCTGGAG CACAGTTAA GAACTCTACGT GCTGTTTCAG
 851 AATCCGGAGT TTATTTCTAT GATCCTATAA GCCATAGCGA GTCGCATAAA
 901 ATTACAGATC TTGTAATCAA TGCTCCGTGAA GGAAAGGAAA CTTATGAAGG
 951 AACAAATTAGC TTCTCAGGAC TATGCCCTGGA TGATCATGAA GTTTGTGCGG
 1001 AAAATCTTAC TTCCACAAATC CTACAAGATG TCACATTAGC AGGAGGAAC
 1051 CTCTCTCTAT CGGATGGGT TACCTTGCAA CTGCATTCTT TTAAGCAGGA
 1101 AGCAAGCTCT ACGCTTACTA TGTCTCCAGG AACCACTCTG CTCTGCTCAG
 1151 GAGATGCTCC GTTCTCAGGAC ATGGCACATCC TGATTGAGA TACCGACAAC
 1201 TTTGTTCCTG TAAGGATTGCG CGCCGAGGAC AAGGATGCTC TTGACTTCATT
 1251 AGAAAAAAACTT AAAGTTGCCT TTGAGGCTTA TTGGTCCGTC TATGACTTT
 1301 CTCAATTAA GGAAGCCTTT ACGATTCCCTC TTCTTGAAC TCTAGGGCCT
 1351 TCTTTTGACA GTCTTCTCCT AGGGGAGACC ACTTTGGAGA GAACCCAAGT
 1401 CACAACAGAG AATGACGCCG TTGAGGTTT CTGGTCCCTA AGCTGGGAAG
 1451 AGTACCCCCC TTCTCTGGAT AAAGACAGAA GGATCACACC AACTAAGAAA
 1501 ACTGTTTCC TCACTTGGAA TCCTGAGATC ACTTCTACGC CATAA

45 The PSORT algorithm predicts an outer membrane location (0.922).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product, as shown in Figure 26A. The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 26B).

These experiments show that cp6735 is a surface-exposed and immunoaccessible protein, and that it 50 is a useful immunogen. These properties are not evident from the sequence alone.

Example 27

The following *C.pneumoniae* protein (PID 4376784) was expressed <SEQ ID 53; cp6784>:

55 1 MNRRKARWVV ALFAMTALIS VGCCPWSQAK SRCSIDKYIP VVNRLLEVC
 51 LPEAENVEDL IESSSAWLT PEERFSGELV SICQVKDEHA FYNDLSSLHM
 101 TQAVPVSYSAT YDCAVVFGGP LPALRQLDF LVREWQRGVR FKKIVFLC
 151 RGRYQSIEEQ EHFFDSRYNP FPTEENWESG NRVTIPSSEEE IAKFVWMQML

-69-

201 LPRAWRDSTS GVRVTFLAK PEENRVVANR KDTLLLFRSY QEAFFGRVLF
 251 VSSQPIFIGLD ACRVGQFFKG ESYDLAGPGF AQGVLKHYWA PRICLHTLAE
 301 WLKETNGCLN ISEGCFG*

A predicted signal peptide is highlighted.

- 5 The cp6784 nucleotide sequence <SEQ ID 54> is:

1 ATGAATAGAA GAAAAGCAAG ATGGGTAGTG GCATTGTTTCG CAATGACGGC
 51 GCTCAATTCT GTTGGGTGTT GTCCTTGGTC ACAAGCGAAA TCAAGATGTT
 101 CTATTGATAA GTATATTCTC GTAGTCATC GTTACTAGA AGTTTGTGGA
 151 CTTCCCTGAAG CTGAGAAATGT TGAGGATTAA ATCGAGTCCT CGTCTGCTTG
 201 GGTACTGACT CCTGAAGAAC GTTTTCTGG AGAGTTAGTC TCTATCTGTC
 251 AGGTTAAAGA TGAGCATGCT TTCTATAAACG ATTGTCTTT ATTACATATG
 301 ACTCAGGCTG TGCCCTCGTA TTCTGCAACG TATGATTGTC CTGTAAGTTT
 351 TGGCGGGCCT TTGCCAGCGC TAGTCAGCG CTTAGATTT TTGGTGCAGAG
 401 AGTGGCAGCG TGGCGTGCAGC TTTAAGAAAAA TCCTTTCTG ATGTGGAGAG
 451 CGAGGGCGCT ATCAGTCAT TGAAGAACAA GAGCATTTCT TTGATTCTCG
 501 GTACAATCTT TTCCCTACTG AAGAGAACATG GGAATCTGGT AACCGAGTTA
 551 CTCCCTCTTC TGAAGAACAG ATTGCCAAT TTGTTGGAT GCAAATGCTT
 601 TTACCTAGAG CATGGCGAGA TAGTACTTCG GGAGTCAGAG TGACATTTCT
 651 TCTAGCAAAG CCAGAGGAAA ATCGTGTGGT TGCGAATCGT AAGGACACCT
 701 TACTTTTATT CCGTTCTTAT CAAGAACCGT TTCCGGGACG CGTGTATTT
 751 GTAAGTAGTC AACCCTTTAT CGGTTTAGAT GCTTGCAGGG TCGGGCAGTT
 801 TTCAAAGGG GAAAGCTATG ATCTTGCTGG ACCTGGATTT GCTCAAGGAG
 851 TCTTGAAAGTA TCATTGGGCT CCAAGGATTT GTCTACATAC TTTAGCGGAA
 901 TGGTTAAAGG AAACGAACGG CTGCTTAAAT ATTCAGAGG GTTGTGTTGG
 951 ATGA

The PSORT algorithm predicts a periplasmic location (0.894).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product, as shown in Figure 27A. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 27B). The GST-fusion product was used for FACS analysis (Figure 27C).

- 30 The cp6784 protein was also identified in the 2D-PAGE experiment (Cpn0498).

These experiments show that cp6784 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 28

The following *C.pneumoniae* protein (PID 4376960) was expressed <SEQ ID 55; cp6960>:

35 1 MNRRWNLVLA TVALALSVAS CDVRSKDKDK DQGSLVEYKD NKDTNDIELS
 51 51 DNQKLSRTFG HLLARQLRKS EDMFFDIAEV AKGLQAEVLC KSAPLTETEY
 101 EEKMAEVQKL VFEKKSKENL SIAEKFLKEN SKNAGVVEVQ PSKLQYKIIK
 151 EGAGKAISGK PSALLHYKGS FINGQVFSSS EGNNEPILLP LGQTIPGFAL
 201 GMQGMKEGET RVLYIHPDLA YGTAGQLPPN SLLIFEINLI QASADEVAAV
 40 251 PQEGQNQE*

A predicted signal peptide is highlighted.

The cp6960 nucleotide sequence <SEQ ID 56> is:

45 1 ATGAACAGAC GGTGGAATTT AGTTTTAGCA ACAGTAGCTC TGGCACTCTC
 51 51 CGTCGCTTCT GTGTGACGTAC GGTCTAAGGA TAAAGACAAAG GATCAGGGGT
 101 101 CGTTAGTGGG ATATAAAAGAT AACAAAGATA CCAATGACAT AGAATTATCC
 151 151 GATAATCAA AGTTATCCAG AACATTTGGT CATTATTTAG CACGCCAATT
 201 201 ACGCAAGTCA GAAGATATGT TTTTTGATAT TGCAGAAAGTG GCTAAGGGGT
 251 251 TGCAGGGGAA ATTGGTTGT AAAAGTGCTC CTTTAACAGA AACAGAGTAT
 301 301 GAAGAAAAAA TGGCTGAAGT ACAGAAAGTTG GTTTTTGAAA AAAAATCAA
 351 351 AGAAAATCTT TCATTGGCAG AAAAATCTT AAAAGAAAAT AGCAAGAACG
 401 401 CTGGTGTGTT TGAAGTCAA CCAAGTAAAT TGCAATACAA AATTATTA

451 GAAGGTGCAG GGAAAGCAAT TTCAAGTAAA CCTTCAGCTC TATTGCACTA
 501 CAAGGGTTCC TTCATCAATG GCCAAGTATT TAGCAGTTCA GAAGGCAACA
 551 ATGAGCCTAT CTTGCTTCCT CTAGGCCAAA CAATTCTGG TTTTGCTTTA
 601 GGTATGCAGG GCATGAAAGA AGGAGAAACT CGAGTTCTCT ACATCCATCC
 651 TGATCTTGCT TACGGAACCG CAGGACAAC TCCCTCAAAC TCTTTATTAA
 701 TTTTGAAAT TAACTTGATT CAGGCTTCAG CAGATGAAGT TGCTGCTGTA
 751 CCCCAAGAAG GAAATCAAGG TGAATGA

The PSORT algorithm predicts periplasmic space location (0.930).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product, as shown in Figure 28A. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 28B) and for FACS analysis (Figure 28C).

The cp6960 protein was also identified in the 2D-PAGE experiment.

These experiments show that cp6960 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

15 Example 29

The following *C.pneumoniae* protein (PID 4376968) was expressed <SEQ ID 57; cp6968>:

1 MKFLLYVPLL LVLVSTGCDA KPVSFEPFSG KLSTQRFEPQ HSAEYFSQG
 51 QEFLKKGNFR KALLCFGIIIT HHFPRDILRN QAQYLIGVCY FTQDHPDLDAD
 101 KAFASYLQLP DAEYSEELFQ MKYAIQAQRFA QGKRKRICRL EGFPKLMNAD
 151 EDALRIYDEI LTAFPSKD LG AQALYSKAAL LIVKNDLTEA TKTLKKLTQ
 201 FPLHILSSEA FVRLSEIYLO QAKKEPHNLQ YLHFAKLNEE AMKKQHPNHP
 251 LNEVVSANVG AMREHYARGL YATGRFYEK KKAEEAANIYY RTAITNYPDT
 301 LLVAKCQKRL DRISKHTS*

A predicted signal peptide is highlighted.

25 The cp6968 nucleotide sequence <SEQ ID 58> is:

1 ATGAAATTC TATTATACGT TCCACTTCCTT CTIGTTCTCG TATCTACGGG
 51 GTGCGATGCA AAACCTGTTT CTTTGAGCC CTTTCAGGA AAGCTTCCA
 101 CCCAGCGTT TGAGCCTCAG CACTCTGCTG AAGAATATT TTCTCAGGG
 151 CAGGAATTCT TAAAAAAAGG AAATTCAGA AAAGCTTAC TATGCTTGG
 201 ATCATTTACG CATCACCTCC CTAGGGACAT CTTGGTAAT CAACGACAGT
 251 ATCTTATAGG AGTCTGTTAC TTCACGCAGG ATCACCCAGA TTTAGCAGAC
 301 AAGGCATTG CATCTTACTT ACAACTTCCT GATGCGGAGT ACTCTGAAGA
 351 GTTGTTCAG ATGAAATATG CGATTGCTCA AAGATTTGCT CAAGGGAAGC
 401 GTAAACGGAT TTGTCGATTA GAGGGCTTCC CAAAACTAAT GAATGCTGAT
 451 GAAGATGCGC TACGCATTTA TGACGAGATT CTAACAGCGT TTCTTAGTAA
 501 AGACTTAGGA GCTCAGGCC TCTATAGTAA AGCTGCGTTA CTTATTGTAA
 551 AAAACGATCT TACAGAAGCC ACCAAAACCT TAAAAAAACT CACGTTACAA
 601 TTTCCTCTAC ATATTTTATC TTCAGAGGCC TTTGTACGTT TATCGGAAAT
 651 CTATTTACAG CAAGCTAAGA AAGAGCCTCA CAATCTTCAA TATCTTCATT
 701 TTGCAAAGCT TAATGAAGAG GCAATGAAAA AGCACCATCC TAACCACATCCT
 751 CTGAATGAGG TTGTTTCTGC TAATGTTGGA GCTATGGGG AACATTATGC
 801 TCGAGGTTTG TATGCCACAG GTCGTTTCTA TGAGAAGAAG AAAAAAGCCG
 851 AGGCTGCGAA TATCTATTAC CGCACTGCGA TTACAAACTA CCCAGACACT
 901 TTATTAGTGG CTAAATGTCA AAAGCGTCTA GATAGAATAT CTAAGCATAAC
 951 TTCCCTAA

The PSORT algorithm predicts an inner membrane location (0.790).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product, as shown in Figure 29A. The recombinant GST-fusion was used to immunise mice, whose sera were used in a Western blot (Figure 29B) and for FACS analysis (Figure 29C).

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6968 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

5 Example 30

The following *C.pneumoniae* protein (PID 4376998) was expressed <SEQ ID 59; cp6998>:

10	1 MKLLKSALL SAAFAGSVGS LQALPVGNPS DPSLLIDGTI WEGAAGDPCD 51 PCATWCDAIS LRAGFYGDYV FDRILKVDAP KTFSMGAKPT GSAAANYTTA 101 VDRPNPAYNK HLHDAEWFNTN AGFIALNIWD RFDVFCTLGA SNGYIRGNST 151 AFNLVGLFGV KGTTVNANEL PNVSLNSNGVV ELYTDTSFSW SVGARGALWE 201 CGCATLGAEF QYAQS PKPKVE ELNVICNVSQ FSVNPKGYK GVAFPLPTDA 251 GVATATGKTS ATINYHEWQV GASLSYRLNS LPVYIGVQWS RATFDADNIR 301 IAQPKLPTAV LNLTAWNPSL LGNATALSTT DSFSDFMQIV SCQINKFKSR 351 KACGVTVGAT LVDADKWSLT AEARLINERA AHVSGQFR*
----	---

15 A predicted signal peptide is highlighted.

The cp6998 nucleotide sequence <SEQ ID 60> is:

20	1 ATGAAAAAAC TCTTAAAGTC GGCCTTATTAA TCCGCCGCAT TTGCTGGTTTC 51 TGTTGGCTCC TTACAAGCCT TGCCCTGAGG GAACCCCTCT GATCCAAGCT 101 TATTAATTGA TGGTACAATA TGGGAAGGTG CTGCAGGAGA TCCTTGCGAT 151 CCTTGCGCTA CTTGGTGCAG CGCTATTAGC TTACGTGCTG GATTITACGG 201 AGACTATGTT TTCGACCGTA TCTTAAAAGT AGATGCACCT AAAACATTTT 251 CTATGGGAGC CAAGCCTACT CGATCCGCTG CTGCAAACTA TACTACTGCC 301 GTAGATAGAC CTAACCCGGC CTACAATAAG CATTACACG ATGCAGAGTG 351 GTTCACTAAT GCAGGCTTCA TTGCGCTTAAA CATTGGGAT CGCTTTGATG 401 TTTTCTGTAC TTTAGGAGCT TCTAATGGTT ACATTAGAGG AAACCTCTACA 451 GCGTTCAATC TCGTTGGTTT ATTCCGGAGTT AAAGGTACTA CTGTAAATGC 501 AAATGAACTA CCAAACGTTT CTTTAAGTAA CGGAGTTGTT GAACTTTACA -551 CAGACACCTC TTTCTCTGG AGCGTAGGCG CTCGTGGAGC CTTATGGAA 601 TGCGGTTGTG CAACTTGGG AGCTGAATTCA CAATATGCAC AGTCCAAACC 651 TAAAGTTGAA GAACTTAATG TGATCTGTAA CGTATCGCAA TTCTCTGTAA 701 ACAAAACCAA GGGCTATAAA GGCCTTGCTT TCCCCTTGCC AACAGACGCT 751 GGCCTAGCAA CAGCTACTGG AACAAAGTC GCGACCATCA ATTATCATGA 801 ATGGCAAGTA GGAGCCTCTC TATCTTACAG ACTAAACTCT TTAGTGCCT 851 ACATTGGAGT ACAATGGTCT CGAGCAACTT TTGATGCTGA TAACATCCGC 901 ATTGCTCAGC CAAAACTACC TACAGCTGTT TAAACTTAA CTGCATGGAA 951 CCCCTCTTTA CTAGGAAATG CCAACAGCACT GTCTACTACT GATTGTTCT 1001 CAGACTTCAT GCAAATTGTT TCCTGTCAAGA TCAACAAGTT TAAATCTAGA 1051 AAAGCTTGTG GAGTTACTGT AGGAGCTACT TTAGTTGATG CTGATAAATG 1101 GTCACTTACT GCAGAAGCTC GTTTAATTAA CGAGAGAGCT GCTCACGTAT 1151 CTGGTCAGTT CAGATTCTAA
----	--

The PSORT algorithm predicts an outer membrane location (0.707).

The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 30A) and as a his-tag product. The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 30B) and for FACS analysis (Figure 30C).

45 The cp6998 protein was also identified in the 2D-PAGE experiment (Cpn0695) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6998 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 31

The following *C.pneumoniae* protein (PID 4377102) was expressed <SEQ ID 61; cp7102>:

5	1 MKHTFTKRVL FFFFFLVIPIP LLLLNLVVGF FSFSAAKANL VQVLHTRATN 51 LSIEFEKKLT IHKLFLDRLA NTLALKSYAS PSAE PYAQAY NEMM ALSNTD
	101 FSLCLIDPFD GSVRTKNP GD PFIRYLKQHP EMKKKLSAAV GKAFLLTIPG 151 KPLLHYLLLV EDVASWDSTT TSGLLVSFYP MSFLQKDLFQ SLHITKGNIC
	201 LVNKYGEVLF CAQDSESSFV FSLDLPNLPQ FQARSPSAIE IEKASGILGG 251 ENLITVSI NK KRYLGLVLNK IPIQGTYTLS LPVPVSDLIQS ALKVPLNICF
10	301 FYVLAFLLMW WIFSKINTKL NKPLQELTFC MEEAWRGHNH VRFEPQPYGY 351 EFNELGNIFN CTLLLLLNSI EKADIDYHSG EKLQKELGIL SSLQSALLSP
	401 DFPTFPKVTF SSQHLRRRQL SGHFNGWTVQ DGGDTLLGII GLAGDIGLPS 451 YLYALSARSRL FLAYASSDVS LQKISKDTAD SFSKTTEGNE AVVAMTFIKY
	501 VEKDRSLELL SLSEGAPTMF LQRGESFVR PLETHQALQP GDRLICLTGG 551 EDILKYFSQL PIEELLKDPL NPLNTENLID SLTMMLNNET EHSADGTLTI
15	601 LSFS*

A predicted signal peptide is highlighted.

The cp7102 nucleotide sequence <SEQ ID 62> is:

20	1 ATGAAACATA CCTTTACCAA GCGTGTCTA TTTTTTTCT TTTTAGTGAT 51 TCCCATTCCC CTACTCCTCA ATCTTATGGT CGTAGGTTTT TTCTCATTTT 101 CTGCCGCTAA AGCAAATTAA GTACAGGTCC TCCATACCCG TGCTACGAAC 151 TTAAGTATAG AATTGAAAAA AAAACTGACG ATACACAAGC TTTTCCTCGA 201 TAGACTTGCC AACACATTAG CCTTAAAATC CTATGCATCT CCTCTGTGAG 251 AGCCCTATGC ACAGGCATAC AATGAGATGA TGGCACTCTC CAATACAGAC 301 TTTTCCTTAT GCCTTATAGA TCCCTTGTAT GGATCTGTAA GGACGAAAAAA 351 TCCTGGAGAC CCTTTCATTC GCTATCTAAA ACAGCATCTT GAAATGAAGA 401 AAAAGCTATC CGCAGCTGTA GGGAAAGCCT TTTTATTGAC CATTCCAGGT 451 AAACCACTTT TACATTATCT TATTCTAGTT GAAGATGTCG CATCTGGGA 501 TTCTACAACG ACTTCAGGAC TGCTTGTAAAG TTTCTATCCC ATGTCCTTTT 551 TACAGAAAGA TTTATTCCAA TCCTTACACA TCACCAAAGG AAATATCTGC 601 CTTGTAATAA AGTATGGCGA GGTCCCTCTTC TGTCCTCAGG ACAGTGAATC 651 TTCTTTGTAA TTTTCTCTAG ATCTCCCTAA TTTACCGAA TTCCAAGCAA 701 GAAGCCCCCTC TGCCATAGAA ATTGAGAAAG CTTCTGGAAT TCTTGGTGGG 751 GAGAACCTAA TCACAGTGAG TATCAACAAG AAACGCTACC TAGGATTGGT 801 ACTGAATAAA ATTCCATATCC AAGGGACCTA CACTCTATCT TTAGTTCCAG 851 TTTCTGATCT CATCCAATCC GCCTTGAAAG TTCCCTCTCAA TATTGTTTT 901 TTCTATGTAC TTGCTTCCCT CCTCATGTGG TGGATTTCT CTAAGATCAA 951 CACCAAACCTT AACAAAGCTC TTCAAGAACT GACCTTCTGT ATGGAAGCTG 1001 CCTGGCGAGG AAACCATAAC GTGAGGTTTG AACCCCAGCC TTACGGTTAT 1051 GAATTCAATG AACTAGGAAA TATTTCAAT TGCACTCTCC TACTCTTATT 1101 GAATTCCATT GAGAAAGCAG ATATCGATTA CCATTCAGGC GAAAAATTAC 1151 AAAAAGAATT AGGGATTTA TCTTCACTAC AAAGTGGCTT ACTAAGTCCG 1201 GATTTCCCTAA CGTCCCCCTAA AGTTACCTTT AGTTCCCAAC ATCTCCGGAG 1251 AAGGCAACTT TCCGGTCATT TTAATGGTTG GACAGTTCAA GATGGTGGCG 1301 ATACCCTTTT AGGGATCATA GGGCTCGCTG GCGATATTGG TCTTCCCTCC 1351 TATCTCTATG CTTTATCCGC ACGGAGTCTT TTTCTTGCCCT ATGCTTCCCTC 1401 GGACGTTTCG TTACAAAAAA TCAGCAAGGA TACTGCCGAC AGCTTCTCAA 1451 AAACAACAGA AGGCAATGAG GCTGTAGTG CTATGACTTT CATTAAATAT 1501 CTAGAAAAAG ATCGATCTCT AGAGCTCCTC TCGTTAACGCG AGGGAGCTCC 1551 TACCATGTTT CTACAAACGAG GAGAACATTTT CGTACGTCCTC CCCTTAGAGA 1601 CTCACCAAGC TCTACAGCCT GGAGATCGGT TGATCTGCCCT CACTGGAGGA 1651 GAAGACATCC TCAAGTACTT TTCTCAGCTT CCTATTGAAG AGCTCTTAAA 1701 AGATCCTTTA AACCCCTCTAA ATACAGAGAA TCTTATTGAT TCTCTAACCA 1751 TGATGTTAAA CAACGAAACC GAACATTCTG CAGATGGAAC TCTGACCAC 1801 CTTTCATTTT CATAA
----	--

55 The PSORT algorithm predicts an inner membrane location (0.338).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product. The purified GST-fusion product is shown in Figure 31A. The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot and for FACS analysis (Figure 31B).

These experiments show that cp7102 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 32

The following *C.pneumoniae* protein (PID 4377106) was expressed <SEQ ID 63; cp7106>:

```

5      1 MKDLGTLGGT SSTAKTVSPD GKVIMGRSQI ADGSWHAFMC HTDFSSNNVL
      51 FDLDNTYKTL RENGRQLNSI FNQNMMQLQR ASDHEFTTEFG RSNIALGAGL
     101 YVNALQNLPS NLAAQYFGIA YKIRPKYRLG VFLDHNFSSH VPNNFNVSHN
     151 RLWMGAFIGW QDSDALGSSV KVSFGYGKQK ATITREQLEN TEAGSGESHF
    201 EGVAAQIEGR YGKSLGGHVR VQPFLGLQFV HITRKEYTEN AVQFPVHYDP
    251 IDYSTGVVYL GIGSHIALVD SLHVGTRMGM EQNFAAH TDR FSGSIASIGN
    301 FVFEKLVDVTH TRAFAEMRVN YELPYLQSLN LILRVNQQPL QGVMGFSSDL
    351 RYALGF*

```

The cp7106 nucleotide sequence <SEQ ID 64> is:

```

15     1 ATGAAAGATT TGGGGACTCT TGGGGGTACC TCTTCTACAG CAAAAACAGT
      51 GTCCCCAGAT GGTAAAGTGA TCATGGTAG ATCACAAATT GCTGATGGCA
     101 GTTGGCACGC ATTTATGTGT CATA CGGATT TCTCCTCTAA TAATGTACTC
     151 TTTGATCTCG ATAATACGTA TAAA ACTCTA AGAGAAAAATG GCCGTCAGCT
     201 AAATTCCATA TTCAACCTAC AAAATATGAT GTTACAGAGA GCCTCAGATC
     251 ATGAGTTCAC AGAGTTTGGG AGGAGTAACA TCGCTCTTGG TGCCGGGCTT
     301 TATGTGAATG CCTTGAGAAA TCTCCCTAGC AATTTAGCAG CACAATATTT
     351 TGGAAATCGCA TACAAAATAC GTCC TAAATAA TCGTTTGGGG GTGTTTTGG
     401 ACCATAATTT CAGCTCCCAC GTTCC TAAATAA ATTTAACGT AAGCCACAAT
     451 AGACTCTGGA TGGGAGCCTT TATTGGATGG CAGGATTCTG ATGCTCTAGG
     501 ATCTAGTGTCA AAGGTGTCTT TCGGATATGG AAAACAAAAA GCCACGATTA
     551 CAAGAGAGCA ATTAGAGAA ACAGAAGCCG GGAGTGGGG AAGCCATTTC
     601 GAAGGGGTCG CTGCTCAGAT AGAAGGGCGG TATGGTAAGA GCCTCGGAGG
     651 ACATGTCAGG GTCCAGCCTT TCCTAGGACT GCAGTTTGT CACATTACAA
     701 GGAAAGAATA TACCGAAAAT GCAGTGCAAT TTCCTGTACA CTATGATCCT
     751 ATAGACTATT CTACAGGTGT AGTGTATTTA GGAATTGGAT CTCATATTGC
     801 ACTTGTAGAT TCTTTACATG TAGGCACACG CATGGGAATG GAGCAAAACT
     851 TTGCGAGCCA TACGGACAGG TTCTCAGGAT CTATAGCGTC TATTGGAAAC
     901 TTTGTGTITG AAAAGCTGA TGTGACTCAC ACAAGGGCAT TTGCGGAAAT
     951 GCGGTGTCAAC TATGAGCTTC CCTATCTACA GTCTCTGAAT CTTATTCTAC
    1001 GAGTTAATCA ACAGCCTCTA CAAGGGGTTA TGGGATTTC CAGTGATCTT
    1051 AGGTATGCCT TAGGATTCTA A

```

The PSORT algorithm predicts a cytoplasmic location (0.224).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product. The purified GST-fusion product is shown in Figure 32A. The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 32B) and for FACS analysis (Figure 32C).

This protein also showed very good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7106 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 33

The following *C.pneumoniae* protein (PID 4377228) was expressed <SEQ ID 65; cp7228>:

```

1      1 MTAVLILTSF PSEESARSLSA RHLITERLAS CVHVFPKGTS TYLWEGLCE
      51 SEEHHIQIKS IDIRFSEICL AIQEFSGYEV PEVLLFPIEN GDPRYLNWL
     101 ILSYPEKPPL SD*

```

The cp7228 nucleotide sequence <SEQ ID 66> is:

```

5      1 ATGACTGCTG TTCTTATTCT TACATCTTC CCTTCGGAGG AAAGTGCCTG
      51 CTCCTTAGCT AGACATCTGA TTACAGAGCG TCTTGCTTCC TGTGTGCATG
     101 TATTCCCTAA AGGCACATCG ACATATCTAT GGGAAAGGCAA GCTATGTGAG
     151 TCTGAAGAAC ATCATATAACA AATCAAATCG ATAGACATAC GCTTCTCGGA
     201 AATTTGTCTT GCTATTCAAG AGTTCTCTGG CTATGAGGTT CCTGAAGTCT
     251 TACTATTTC TATTGAAAAT GGGGATCCGA GGTACTTGAA TTGGTTAACG
     301 ATTCTCAGCT ATCCAGAGAA GCCTCCGCTT TCAGATTAG

```

The PSORT algorithm predicts an inner membrane location (0.040).

- 10 The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product, as shown in Figure 33A (his-tag = left-hand arrow, GST = right-hand arrow). The proteins were used to immunise mice, whose sera were used in a Western blot (Figure 33B) and FACS analysis.

These experiments show that cp7228 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

15 Example 34

The following *C.pneumoniae* protein (PID 4377170) was expressed <SEQ ID 67; cp7170>:

```

20      1 MNSKMLKHLL LATLSFSMFF GIVSSPAVYA LGAGNPAAPV LPGVNPEQTG
      51 WCAFQLCNST DLFAALAGSL KFGFYGDYVF SESAHITNVP VITSVTTSGT
     101 GTTPPTITSTT KNVDFDLNNNS SISSSCVFAT IALQETSPAA IPLLIDIAFTA
     151 RVGGGLKQYYR LPLNAYRDFT SNPLNAESEV TDGLIEVQSD YGIVWGLSLQ
     201 KVLWKDGVSF VGVSADYRHG SSPINYIIVY NKKANPEIYFD ATDGNLNSYKE
     251 WSASIGISTY LNDYVLPYAS VSIGNTSRKA PSDSFTTELEK QFTNFKFIR
     301 KITNFDRVNF CFGTTCCISN NFYYSVGRW GYQRRAINITS GLQF*

```

A predicted signal peptide is highlighted.

- 25 The cp7170 nucleotide sequence <SEQ ID 68> is:

```

30      1 ATGAATAGCA AGATGCTAAA ACATTTACGT TTAGCCAACCC TTTCCCTCTC
      51 TATGTTCTTC GGGATTGTAT CTTCTCCCGC AGTATATGCC CTAGGGGCTG
     101 GAAACCCCTGC AGCTCCAGTA CTCCCAGGTG TGAATCCTGA GCAAACGGGA
     151 TGGTGTGCCT TCCAACCTTG TAATAGTTAC GATCTTTTG CTGCTCTTGC
     201 AGGAAGCCTC AAATTTGGGT TCTATGGAGA TTATGTCCTC TCAGAAAAGTG
     251 CCCATATTAC CAATGTCCT GTCATTAACCT CCGTTACGAC TTCAGGCACA
     301 GGAACAAACGC CAACCATTAC CTCTACAAC TAAAACGTAG ACTTTGATCT
     351 TAACAAACAGC TCCATCAGCT CGAGCTGTGT TTTTGCAACC ATAGCTCTAC
     401 AGGAAACATC CCCAGCTGCC ATTCCCCCTT TAGATATAGC CTTCACTGCA
     451 CGTGTCCGGAG GACTTAAGCA GTACTACCAGC CTCCTCTCA ATGCTTACAG
     501 AGACTTCACT TCAAATCCTT TAAATGCAGA ATCTGAAGTT ACAGATGGTC
     551 TCATTGAAGT CCAGTCAGAC TATGGAATTG TCTGGGGTCT GAGTTTACAA
     601 AAAGTATTGT GGAAAGATGG AGTGTCTTT GTAGGGGTGA GCGCTGACTA
     651 CCGTCACGGT TCCAGTCCCA TCAACTATAT CATCGTTAC AACAAAGCCA
     701 ACCCCGAGAT CTATTCGAT GCTACTGATG GAAACCTAAG CTATAAAGAA
     751 TGGTCTGCAA GCATCGGCAT CTCTACGTAT CTTAATGACT ATGTGCTTCC
     801 CTATGCATCC GTATCTATAG GAAATACTTC AAGAAAAAGCT CCTTCTGATA
     851 GCTTCACAGA ACTCGAAAAG CAATTACGA ATTTTAAATT TAAAATTGCT
     901 AAAATCACAA ACTTCGACAG AGTAAACTTC TGCTTCGGAA CTACCTGCTG
     951 CATCTCAAAT AACTTCTACT ATAGTGTAGA AGGCCGTTGG GGATATCAGC
    1001 GTGCTATCAA CATTACGTCA GGTCTGCAGT TTTAG

```

The PSORT algorithm predicts a bacterial outer membrane location (0.936).

- The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product. The purified GST-fusion product is shown in Figure 34A. The GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (34B) and for FACS analysis (34C).

The cp7170 protein was also identified in the 2D-PAGE experiment (Cpn0854).

These experiments show that cp7170 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 35

- 5 The following *C.pneumoniae* protein (PID 4377072) was expressed <SEQ ID 69; cp7072>:

```

1  MDIKLFLCLF LCSSLIAMSP IYGKTDYEK LTLTGINIID RNGLSETICS
51 KEKLKYTKV DFLAPQPYQK VMRMYKNKRG DNVSLTAYH TNGQIKQYLE
101 CLNNRAYGRY REWHVNGNIK IQAEVIGGIA DLHPSAESGW LFDQTTFAYN
151 DEGILEAAIV YEKGLLEGSS VYYHTNGNIW KECPYHKGP QGKFLTYTSS
201 GKLLKEQNYQ QGKRHGLSIR YSEDSEEDVL AWEEYHEGRL LKAELYDPQT
251 HEIYATIHEG NGIQAIYGYK AVIETRAFYR GEPYGVKTRF DNSGTQIVQT
301 YNLLQGAKHG EEEFFYPETG KPKLLLWHE GILNGIVKTW YPGGTLESCK
351 ELVNNKKSGL LTIYYPEGQI MATEEYDNDL LIKGEYFRPG DRHPYSKIDR
401 GCGTAVFFSS AGTITKKIPY QDGKPLLN*

```

- 15 A predicted signal peptide is highlighted.

The cp7072 nucleotide sequence <SEQ ID 70> is:

```

1  ATGGATATAA AAAAACCTCTT TTGCTTATTT CTATGTTCTT CTCTAATTGC
51  CATGAGTCCC ATTATGGGA AAACAGGTGA CTATGAGAAA CTCACCCCTTA
101 CAGGGATCAA TATCATTGAT AGAAACGGCC TGTCAGAAAC TATTTGCTCT
151 AAAGAGAACG TAAAGAAATA CACCAAGGTA GACTTCTTG CTCCCCAGCC
201 CTATCAAAG GTCATGAGGA TGTAAAAAA CAAACGCGGA GATAACGTTT
251 CTTGTTAAC AGCCTATCAC ACTAACGGGC AAATTAAAGCA GTACCTGGAG
301 TGTCATCAAATCCTTA TGGAAGATAT CGTGAATGGC ACGTCAACGG
351 GAATATCAAATCCTGA AGGTTATCGG AGGTATTGCG GATCTTCATC
401 CCTCAGCAGA GTCTGGCTGG CTATTTGATC AACTACATT TGCCTATAAT
451 GATGAAGGTA TCTTAGAAGC CGCTATCGTC TATGAAAAAG GGCTGCTCGA
501 AGGATCTTCG GTGTATTACC ATACTAATGG GAATATTGG AAAGAGTGTG
551 CCTATCATAA GGGAGTTCCCT CAAGGTAAT TCCTGACATA CACATCTTCG
601 GGGAAACTGC TCAAAGAACAA GAATTACCAA CAAGGCAAA GACACGGTCT
651 TTGATTTCGC TACAGCGAAG ATTCCGAAGA AGATGTTTTA GCCTGGGAAG
701 AATATCATGA GGGACGACTC CTAAAAGCAG AGTACTTACA TCCTCAAAC
751 CACGAAATCT ATGCGACTAT ACACGAAGGG AACGGCATTG AAGCAATCTA
801 CGGCAAGTAT GCGTTATAG AACTAGGGC ATTTTACCGA GGGGAACCTT
851 ATGGAAAAGT TACCAAGATTC GACAACCTCG GAACACAGAT TGTCCAAACG
901 TATAACCTTT TGCAAGGCAGC GAAGCACCGA GAAGAATTTC TCTTTTATCC
951 TGAGACAGGG AAACCCAAGC TGCTTCTTAA TTGGCATGAA GGAATTTTAA
1001 ATGGGATAGT AAAAACCTGG TATCCCGGAG GAACCTTACA AAGTTGTAAA
1051 GAACTCGTAA ATAACAAAAA ATCCGGTTA CTGACCATT ACTACCCTGA
1101 AGGACAGATC ATGGCGACCG AAGAGTATGA TAATGATCTT CTAATTAAAG
1151 GAGAGTACTT CCGCCCTGGA GACCGTCATC CCTACTCTAA AATAGATCGT
1201 GGTGTGGGA CTGCACTATT TTTCTCGTCG GCGGGAACTA TTACTAAAAA
1251 AATCCCCTAT CAGGACGGCA AACCTTTGCT CAACTAG

```

The PSORT algorithm predicts a periplasmic location (0.688).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 35A) and as a GST-fusion product (Figure 35B). The recombinant his-tag protein was used to immunise mice, whose sera were used in a Western blot (Figure 35C) and for FACS analysis.

These experiments show that cp7072 is a useful immunogen. These properties are not evident from the sequence alone.

Example 36

- 50 The following *C.pneumoniae* protein (PID 4376879) was expressed <SEQ ID 71; cp6879>:

5 1 MATPAQKSPT FQDPSFVREL GSHPVFSPL TLEERGEMAI ARVQQCGWNH
 51 TIVKVSLIIL ALLTILGGGL LVGLLPAVPM FIGTGLIALG AVIFALALIL
 101 CLYDSQGLPE ELPPVPEPQQ IQIEDLRNET REVLEGTLLE VLLKDRDAKD
 151 PAVPQVVVDC EKRLGMLDRK LRREEEILYR STAHLKDEER YEFLLLELEM
 201 RSLVADRLEF NRRSYERFVQ GIMTVRSEEG EKEISRLQDL ISLQQQTVDQ
 251 LRSRIDDEQK RCWTALQRIN QSQKDIQRAH DREASQRACE GTEMIDCAERQ
 301 QLEKDLRQL KSMQEWEIEMR GTIHQQEKAW RKQNAKLERL QEDLRLTGIA
 351 FDEQSLFYRE YKEKYLSQKL DMQKILQEVN AEKSEKACLE SLVHDYEKQL
 401 EQKDANLKKA AAVWEEEELGK QQQEDYEQTQ EIRRLSTFIL EYQDSLREAE
 451 KVEKDFQELQ QRYSRQLEEK QVKKEKILEES MNHFADLFK AQKENMAYKK
 501 KLADELEGAAA PTEIGEDDDW VLTDASLSQ KKIRELVEEN QELLKALAFK
 551 SNELTQLVAD AVEAEKEISK LREHIEEQKE GLRALDKMHA QAIKDCEAAQ
 601 RKCCDLESLL SPVREDAGMR FELEVELQRL QEENAQLRAE VERLEQEQQFQ
 651 G*

15 The cp6879 nucleotide sequence <SEQ ID 72> is:

20 1 ATGGCAACAC CCGCTCAAAA ATCCCCTACA TTTCAAGATC CTAGTTTTGT
 51 AAGAGAGCTA GGCAGTAACC ACCCTGTCTT TTCCCCGCTA ACGCTTGAGG
 101 AAAGAGGGGA GATGGCAATA GCTCGAGTC AGCAGTGTGG ATGGAATCAT
 151 ACAATTGTTA AGGTAAGTCT TATTATTCTT GCTCTTCTTA CTATTTTAGG
 201 GGGAGGATTA CTCGTAGGAT TGCTGCCAGC AGTTCCATG TTTATTGGAA
 251 CAGGTCTGAT TGCTTTGGGA GCCGTTATAT TTGCTTTGGC TTTGATTTTA
 301 TGCTTTATG ATTCTCAGGG CCTTCCTGAG GAACTCCCTC CGGTTCTGA
 351 ACCACAACAA ATTCAAGATTG AAGATTTAAG AAACGAGACC AGAGAAGTTC
 401 TTGAAGGGAC TCTTTTAGAG GTTCTCTTAA AGGATAGAGA CGCTAAGGAC
 451 CCTGCGGTGC CCCAGGTGGT TGTAGACTGT GAAAAGCGTC TTGGAATGTT
 501 GGATCGTAAG CTGCGACGTG AAGAGGAGAT TCTGTATCGC TCGACGGCCC
 551 ATCTAAAGA CGAGGAAAGG TATGAGTTCT TGCTGGAGCT CTTGGAAATG
 601 CGTAGTCTGG TTGCGGATCG GCTAGAATT AACCCTAGAA GTTATGAGCG
 651 ATTGTTCAA GGAATTATGA CAGTTAGATC AGAGGAGGG GAAAAAGAGA
 701 TTCTCGTCT ACAAGATCTA ATCAGTTGC ACCAGCAGAC GGTGCAAGAT
 751 TTAAGGAGTC GGATCGATGA CGAGCAGAAC AGATGCTGGA CGGCTTTACA
 801 ACGTATTAAAC CAATCTCAGA AGGATATACA ACGGGCTCAT GATCGCGAGG
 851 CTCGCAGCG TGCCTGTGAG GGCACAGAGA TGGATTGTGC AGAACGCCAG
 901 CAACTGGAGA AGGATTTAAG GAGACAGCTG AAATCTATGC AGGAGTGGAT
 951 TGAGATGAGG GGCACAATCC ATCAACAAGA GAAGGCTTGG CGTAAGCAGA
 1001 ATGCCAAATT AGAAAGATTA CAAGAGGATC TGAGACTTAC TGGGATTGCT
 1051 TTTGACGAAC AATCTCTGTT CTATCGCGAA TATAAAAGAGA AATATCTGAG
 1101 TCAGAAACTA GATATGCAAAG AGATTTTACA GGAAGTCAAC GCAGAGAAAA
 1151 GTGAGAAGGC TTGCTTAGAG AGTCTGGTCC ATGACTATGA GAAGCAGCTC
 1201 GAAACAAAAG ATGCTAATCT GAAGAAAGCA GCAGCTGTTT GGGAAAGAAGA
 1251 ATTAGGGAAAG CAGCAACAGG AAGACTACGA ACAAAACCAA GAAATTAGAC
 1301 GTCTGAGTAC ATTCAATTCTT GAGTACCAAGG ACAGTCTGCC TGAGGCAGAA
 1351 AAAGTTGAGA AAGATTCCA AGAGCTACAA CAAAGGTATA GCCGTCTCA
 1401 AGAGGAGAAA CAGGTAAAAG AAAAATCTT AGAAGAAAGT ATGAATCATT
 1451 TTGCGGATCT CTTTGAGAAG GCTCAAAAGG AAAACATGGC CTACAAGAAG
 1501 AAGTTAGAGG .ATTTAGAGGG TGCCGCTGCT CCTACTGAGA TCGGTGAGGA
 1551 CGATGACTGG GTACTCACAG ATTCTGCTTC TCTCAGCCAG AAGAAGATCC
 1601 GCGAACTCGT GGAAGAGAAAT CAAGAACTCC TGAAAGCACT TGCATTTAAA
 1651 TCTAACGAAT TGACTCAACT GGTGCGGAT GCTGTAGAAG CTGAAAAGA
 1701 AATCAGCAAG CTTCGAGAAC ACATAGAAGA GCAGAAAGAA GGATTACGAG
 1751 CTTTGATAA GATGCATGCA CAAGCGATCA AAGATTGCGA AGCTGCTCAG
 1801 AGAAAATGCT GTGACCTTGA GAGCCTTCTC TCTCCTGTTG GAGAAGATGC
 1851 TGGAAATGAGA TTTGAGCTAG AGGTCGAGCT TCAAAGATTG CAAGAAGAAA
 1901 ATGCACAGCT TAGAGCGGAG GTTGAAGAC TAGAGCAAGA GCAATTCAA
 1951 GGATAA

The PSORT algorithm predicts an inner membrane location (0.646).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product.

The purified GST-fusion product is shown in Figure 36A. The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 36B) and for FACS analysis.

60 These experiments show that cp6879 is useful immunogen. These properties are not evident from the sequence alone.

Example 37

The following *C.pneumoniae* protein (PID 4376767) was expressed <SEQ ID 73; cp6767>:

```

5   1 MIKQIGRFFR AFIFIMPLSL TSCESKIDRN RIWIVGNTNAT YPPFEYVDAQ
    51 GEVVGFIDL AKAISEKLGK QLEVREFAFD ALILNLKKHR IDAILAGMSI
   101 TPSRQEIAL LPYYGDEVQE LMVVSRSLE TPVLPLTQYS SVAVQTGTFQ
   151 EHYLLSQPGI CVRSFDSTLE VIMEVRYGKS PVAVLEPSVG RVVLKDFPNL
   201 VATRLELPPE CWVLGCGLGV AKDRPEEIQT IQQAITDLKS EGVIQSLTKK
   251 WQLSEVAYE*

```

The cp6767 nucleotide sequence <SEQ ID 74> is:

```

10  1 ATGATAAAAC AAATAGGCCG TTTTTTTAGA GCATTTATTT TTATAATGCC
    51 TTATCTTTA ACAAGTTGT AGTCTAAAAT CGATCGAAAT CGCATCTGGA
   101 TTGTTAGGTAC GAATGCTACA TATCCTCCTT TTGAGTATGT GGATGCTCAG
   151 GGGGAAGTTG TAGGTTTCGA TATAGATTG GCAAAGGCAA TTAGTGAAAAA
   201 ACTTGGCAAG CAATTGGAAG TTAGAGAATT CGCTTTCGAT GCTTTAATTT
   251 TAAATTTAAA AAAACATCGT ATCGATGCAA TTTTAGCAGG AATGTCCATT
   301 ACTCCTTCGC GTCAGAAGGA AATCGCCCTG CTTCCCTATT ATGGCGATGA
   351 GGTTCAAGAG CTGATGGTGG TTTCTAAGCG GTCTTAGAG ACCCCTGTGC
   401 TTCCCCCTAAC ACAGTATTCT TCTGTTGCTG TTCAGACAGG AACGTTTCAG
   451 GAGCATTATC TTTTATCTCA GCCCGGAATT TGTGTCGTT CTTTTGATAG
   501 CACCTTGGAG GTGATTATGG AAGTCGTTA TGGGAAATCT CCGGTTGCCG
   551 TTCTAGAACCC CTCGGTAGGA CGTGTGTTTC TTAAAGACTT CCCTAATCTT
   601 GTGCAACAA GATTAGAGCT CCCTCCTGAA TGTGGGGTGT TGGGCTGTGG
   651 TCTCGGGCGTA GCTAAAGATC GTCCCTGAAGA AATACAAACG ATTCAACAAG
   701 CGATTACAGA TTTAAAGAGC GAAGGGGTGA TTCAATCTT AACCAAGAAA
   751 TGGCAACTTT CTGAAGTTGC TTACGAATAG

```

The PSORT algorithm predicts an inner membrane location (0.083).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product. The purified his-tag product is shown in Figure 37A. The recombinant his-tag protein was used to immunise mice, whose sera were used in a Western blot (Figure 37B) and for FACS analysis (Figure 37C). The GST-fusion was also used in a Western blot (Figure 37D).

The cp6767 protein was also identified in the 2D-PAGE experiment and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6767 is a useful immunogen. These properties are not evident from the sequence alone.

Example 38

The following *C.pneumoniae* protein (PID 4376717) was expressed <SEQ ID 75; cp6717>:

```

40  1 MMSRLRFRLA ALGIFFILLV PNSVSAKTIV ASDKEKVGVL VYDNSVEAFQ
    51 QILDCIDHAN FYVELCPCMT GGRTLKEMVD HLEARMDLVP ELCSYIIIQP
   101 TFTDAEDQKL LKALKERHPN RFFYVFTGCP PSTSILAPNV IEMHIKLSII
   151 DGKYCILGGT NFEEFMCTPG DEVPEKVDNP RLFVSGVRRP LAFRDQDIML
   201 RSTAFGLQLR EYHKQFAMW DYYAHMMWF DNPEQFAGAC PPLTLEQAE
   251 TVFPGFDFKHE DLVLVDSKKI RIVLGGPHDK QPNPVTQEYL KLIQGARSSV
   301 KLAHMYFIPK DELLNALVDV SHNHGVHLSL ITNGCHELSP AITGPYAWGN
   351 RINYFALLYG KRYPLWKKWF CEKLKPYERV SIYEFIAIWET QLHKKCMIID
   401 DEIFVIGSYN FGKKSDAFDY ESIVVIESPE VAAKANKVFN KDIGLSIPVS
   451 HGDIFSWYFH SVHHTLGHHLQ LTYPMA*

```

A predicted signal peptide is highlighted.

The cp6717 nucleotide sequence <SEQ ID 76> is:

5 1 ATGATGAGTC GGTTGCCTT TCGCTTGGCA GCTCTGGAA TATTTTTAT
 51 TTTGCTGGTT CCTAATTCTG TTTCAGCAAA GACAATCGTA GCTTCAGACA
 101 AGGAGAAGGT TGGAGTTCTT GTTATGACA ATAGTGTAGA GGCCTTCAA
 151 CAGATATTGG ATTGCATAGA TCATGCAAAT TTTTATGTAG AACTGTGTCC
 201 CTGCATGACA GGAGGCCAA CGCTTAAAGA GATGGTAGAT CACCTCGAGG
 251 CTCGTATGGA TCTGGTCCA GAGCTCTGTA GCTATATCAT TATCCAACCC
 301 ACGTTTACCG ATGCTGAAGA CCAAAATTA CTCAAAGCTC TCAAAGAACG
 351 TCATCCCAAC CGGTTTCT ACGBTTTAC AGGGTGCCCA CCCTCAACAA
 401 GCATCCTCGC TCCTAATGTC ATTGAAATGC ATATCAAAT TTCTATCATC
 451 GATGGGAAAT ATTGTATTT AGGTGGTACC AATTTGAAG AGTTTATGTG
 501 CACTCCAGGG GATGAGGTT CTGAGAAAGT GGATAACCCA CGTTTATTG
 551 TCACTGGAGT GCGTCGGCCC CTAGCATTTG GTGATCAGGA TATCATGTTG
 601 CGTTCTACAG CATTGGTTT GCAGCTCAGA GAAGAATATC ATAAGCAATT
 651 TGCTATGTGG GACTACTATG CACATCATAT GTGGTTCATT GATAATCCTG
 701 AACAGTTGC AGGCCCTGT CCTTCACTGA CTTTAGAACAA AGCCGAGGAG
 751 ACAGTATTTC CTGGATTGAA CAAACATGAA GATCTTGTTC TTGTCGACTC
 801 TTCCAAGATC AGGATAGTT TAGGTGGTCC CCACGATAAG CAACCCAAATC
 851 CTGTGACTCA AGAATATTG AAACCTTATCC AGGGAGCTAG ATCTTCTGTG
 901 AAGCTTGCTC ACATGTATTT CATCCCTAAAG GACGAGCTT TAAATGCTCT
 951 TGTGACGTT TCTCATATAAC ACGGTGTCA TCTGAGTTA ATTACGAACG
 1001 GCTGTCATGA ATTAAGTCTT CGAACATTACAG GACCCTATGC TTGGGGAAAC
 1051 CGTATTAACAT ATTCGCCTT GCTCTATGGG AAACGGTATC CTCTTGGAA
 1101 AAAATGGTTT TGCAGAAAGC TAAAACCTTA TGAGCGGGTT TCTATTATG
 1151 AGTTTGCTAT TTGGGAAACG CAGTTGCACA AGAAGTGTAT GATTATCGAT
 1201 GATGAAATTG TTGTGATCGG AAGTTATAAT TTTGGAAAGA AAAGTGTGATC
 1251 CTTGATTAC GAAAGTATTG TAGTTATCGA ATCTCCAGAA GTCGCTGCAA
 1301 AAGCTAACAA AGTCTTCAAT AAAGATATCG GATGTCGAT TCCTGTAAGT
 1351 CATGGCACA TTTCTCTTG GTATTTCAT TCCGTACACC ACACTTGGG
 1401 ACATTGCACTGACCTATA TGCCAGCCTA G

30 The PSORT algorithm predicts a periplasmic location (0.939).

The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 38A), as a his-tagged protein, and as a GST/his fusion product. The proteins were used to immunise mice, whose sera were used in a Western blot (Figure 38B) and for FACS analysis.

These experiments show that cp6717 is a useful immunogen. These properties are not evident from
 35 the sequence alone.

Example 39

The following *C.pneumoniae* protein (PID 4376577) was expressed <SEQ ID 77; cp6577>:

40 1 MKKLLFSTFL LVLGSTSAAH ANLGYVNLKR CLEESDLGKK ETEELEAMKQ
 51 QFVKNAEKIE EELTSIYNLQ QDEDYMESLS DSASEELRKK FEDLSGEYNA
 101 YQSQYYQSIN QSNVKRIQKL IQEVKIAAES VRSKEKLEAI LNEEAFLAIA
 151 PGTDKTTEII AILNESFKKQ N*

A predicted signal peptide is highlighted.

The cp6577 nucleotide sequence <SEQ ID 78> is:

45 1 ATGAAAAAAT TATTATTTTC TACATTTCTT CTGTTTTAG GATCAACAAG
 51 CGCAGCTCAT GCAAATTAG GCTATGTTAA TTTAAAGCGA TGTCTTGAAG
 101 AATCCGATCT AGGTAAAAAG GAAACTGAAG AATTGGAAGC TATGAAACAG
 151 CAGTTGTAA AAAATGCTGA GAAAATAGAA GAAGAACTCA CTTCTATTAA
 201 TAATAAGTTG CAAGATGAAG ATTACATGGA AAGCCTATCG GATTCTGCCT
 251 CTGAAGAGTT GCGAAAGAAA TTCAAGATC TTTCAAGGAGA GTACAATGCG
 301 TACCACTCTC AGTACTATCA ATCTATCAAT CAAAGTAATG TAAAACGCAT
 351 TCAAAAACCT ATTCAAGAAG TAAAATAGC TGCAAGATCA GTGCGGTCCA
 401 AAGAAAAACT AGAAGCTATC CTTAATGAAG AAGCTGTCTT AGCAATAGCA
 451 CCTGGGACTG ATAAAACAC CGAAATTATT GCTATTCTTA ACGAATCTTT
 501 CAAAAAACAA AACTAG

55 The PSORT algorithm predicts a periplasmic space location (0.932).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 39A) and as a GST-fusion product (Figure 39B). The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 39C) and for FACS analysis.

The cp6577 protein was also identified in the 2D-PAGE experiment.

- 5 These experiments show that cp6577 is a useful immunogen. These properties are not evident from the sequence alone.

Example 40

The following *C.pneumoniae* protein (PID 4376446) was expressed <SEQ ID 79; cp6446>:

```

1  MKQPMMSLIFS SVCLGLGLGS LSSCNQKPSW NYHNTSTSEE FFVHGNGKSVS
51  QLPHYPSAFR TTQIFSEEHN DPYVVAKTDE ESRKIWREIH KNLKIKGSYI
101  PISTYGSLMH PKSAALTAKT YRPHPIWING YERSFNIDTG KYLKNGSRRR
151  TSHDGPKNRA VLNLIKSSGR RCNAIGLEMT EEDFVIARR EGVYSLYPVE
201  VCSPYPOGNPF VIAVAYWIADE SACSKEVLPV KGYYSLVWES VSSSDSLNAF
251  GDSFAEDYLR STFLANGTSI LCVHESYKKV PPQP*
```

- 15 A predicted signal peptide is highlighted.

The cp6446 nucleotide sequence <SEQ ID 80> is:

```

1  ATGAAACAGC CCATGTCTCT TATCTTTCA AGTGTATGTT TAGGATTAGG
51  TCTTGGATCT CTTTCCTCCT GTAATCAAAA GCCCTCTTGG AATTATCACA
101  ACACCTAAC GAGCGAAGAAA TTCTTTGTT ATGGAAATAA GAGTGTTCG
151  CAACTGCCTC ATTATCCCTC TGCAATTTCGT ACGACTCAA TCTTTCTGA
201  AGAGCACAAT GATCCTTATG TCGTAGCTAA GACTGATGAA GAGTCTCGTA
251  AAATTTGGAG AGAAATCCAT AAAATCTCA AAATCAAAGG TTCTTACATT
301  CCCATATCGA CTTATGGAAG TCTGATGCAC CCAAAATCAG CAGCTCTTAC
351  ATTAAAAACG TATCGTCCAC ATCCTATTG GATAAAATGGA TACGAGCGTT
401  CTTTTAATAT AGACACAGGA AAGTACTTAA AAAACGGAAG TCGCCGTAGA
451  ACTTCTCACG ATGGTCCGAA AAATCGAGCT GTACTGAATC TCATTAATC
501  TTCGGGACGA CGCTGTAATG CTATAGGCCT TGAGATGACA GAAGAAGACT
551  TTGTAATAGC TAGAAGGCGA GAAGGTGTTT ATAGCCTGTA TCCCCTTGAA
601  GTGTGCTCGT ATCCCTCAGGG GAATCCTTTT GTCATTGCTT ATGCCCTGGAT
651  TGCAGATGAG AGTGCTTGCT CAAAGAGGT CCTACCTGTA AAAGGGTACT
701  ATCTTTAGT CTGGGAAAGC GTTTCTTCCT CTGATTCTCT GAATGCTTTT
751  GGAGATTCCCT TTGCAGAGGA CTACCTCAGA AGCACGTTT TAGCAAACGG
801  AACTTCTATA CTCTGTGTT ATGAAAGCTA TAAGAAAGTT CCTCCTCAGC
851  CCTAA
```

- 35 The PSORT algorithm predicts an inner membrane location (0.177).

The protein was expressed in *E.coli* and purified as a his-tag product and a GST-fusion product. The GST-fusion product is shown in Figure 40A. The recombinant his-tag protein was used to immunise mice, whose sera were used in a Western blot (Figure 40B) and for FACS analysis.

- 40 These experiments show that cp6446 is a useful immunogen. These properties are not evident from the sequence alone.

Example 41

The following *C.pneumoniae* protein (PID 4377108) was expressed <SEQ ID 81; cp7108>:

```

1  MSKKIKVLGH LTLCTLFRGV LCAAALSNIG YASTSQESPY QKSIEDWKGY
51  TFTDLELLSK EGWSEAHAVS GNNSRIVGAS GAGQGSVTAV IWESHLIKHL
101  GTLGGEASSA EGISKDGEVV VGWSDTREGY THAFVFDGRD MKDLGTLGAT
151  YSVARGVSGD GSIIIVGVSAT ARGEDYGWQV GVVKWEKGKIK QLKLLPQGLW
```

201 SEANAISEDG TVIVGRGEIS RNHIVAVKWN KNAVYSLGTL GGSVASAEAI
 251 SANGKIVVGW STTNNGETHA FMHKDETMHD LGTLGGGF SV ATGVSA DGRA
 301 IVGFSAVKTG EIHAFYYAEG EMEDLTTLGG EEARVFDI SS EGNDIIGSIK
 351 TDAGAERAYL FHIHK*

- 5 A predicted signal peptide is highlighted.

The cp7108 nucleotide sequence <SEQ ID 82> is:

1 ATGAGTAAGA AGATAAAAGGT TCTAGGTCAT TTGACGCTCT GCACTCTGTT
 51 TAGAGGAGTG CTGTGTCAG CGGCCCTTTC CAACATAGGA TATGCGAGTA
 101 CTTCTCAGGA ATCACCATAT CAGAAAGTCTA TAGAAGACTG GAAAGGGTAT
 151 ACCTTTACAG ATCTTGAGTT ACTGAGTAAG GAAGGGTGGT CTGAAGCTCA
 201 TGCAGTTCTC GGAAATGGCA GTAGAATTGT AGGAGCTTCG GGAGCTGGCC
 251 AAGGTAGTGT GACTGCTGTC ATATGGGAAA GTCAACCTGAT AAAACATCTC
 301 GGCACCTTCTAG GTGGCGAGGC TTCACTTGCA GAGGGAAATT CAAAGGATGG
 351 AGAGGTGGTC GTTGGGTGGT CAGATACTAG AGAGGGATAT ACTCATGCCT
 401 TTGTCCTTCGA CGGTAGAGAT ATGAAAGATC TCCTGACTCT AGGAGCTACC
 451 TATTCTGTAG CAAGGGGTGT TTCTGGAGAT GGTAGTATCA TCGTAGGAGT
 501 CTCTGCAACT GCTCGTGGAG AGGATTACGG ATGGCAAGTT GGTGTCAAGT
 551 GGGAAAAAGG GAAAATCAA CAATTGAAGT TGTTGCCTCA AGGTCCTCTGG
 601 TCTGAGGCGA ATGCAATCTC TGAGGGATGGT ACCTGATGGT TCGGGAGAGG
 651 GGAAATCTCT CGCAATCACA TCGTTGCTGT AAAATGGAAT AAAATGCTG
 701 TGTATAGTTT GGGGACTCTC GGAGGTAGTG TCGCTTCAGC AGAGGCTATA
 751 TCGGCAAATG GGAAAGTAAT TGTAGGATGG TCCACGACTA ATAATGGTGA
 801 GACTCATGCC TTTATGCACA AAGATGAGAC AATGCACGAT CTCGGCACTC
 851 TAGGAGGAGG TTTTCTGTGTC GCAACTGGAG TTTCTGCTGA TGGGAGAGCC
 901 ATCGTAGGAT TTTCAGCAGT GAAGACCGGA GAAATTCTG CTTTTTACTA
 951 TGCAGAAGGA GAAATGGAGG ATTTAACAAAC TTTGGGAGGG GAAGAAGCTC
 1001 GAGTGTTCGA CATATCTAGC GAAGGAAACG ATATCATTGG CTCTATAAAA
 1051 ACTGACGCTG GAGCTGAACG CGCCTATCTG TTCCATATAC ATAAATAA

The PSORT algorithm predicts an outer membrane location (0.921).

- 30 The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 41A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 41B) and for FACS analysis (Figure 41C). A his-tagged protein was also expressed.

The cp7108 protein was also identified in the 2D-PAGE experiment.

- 35 These experiments show that cp7108 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 42

The following *C.pneumoniae* protein (PID 4377287) was expressed <SEQ ID 83; cp7287>:

1 **MVAKKTVRSY RSSFSHSVIV AILSAGIAFE AKSLHSSELD LGVFNKQFEE**
 51 HSAHVEEAQT SVLKGSVPN PSQKESEKVL YTQVPLTQGS SGESLDLADA
 101 NFLEHFQHLF EETTVFGIDQ KLWSDLDTR NFSQPTQEPD TSNAVSEKIS
 151 SDTKENRKDL ETEDPSKKSG LKEVSSDLPK SPETAVAAIS EDLEISENIS
 201 ARDPLQGLAF FYKNTSSQSI SEKDSSFQGI IFSGSGANS LGFENLKAPK
 251 SGAAVYSDRD IVFENLVKGL SFISCESLED GSAAGVNIVV THCGDVTLTD
 301 CATGLDLEAL RLVKDFSRGG AVFTARNHEV QNNILAGGILS VVGNGKAIIVV
 351 EKNSAEKSNG GAFACGSFVY SNNENTALWK ENQALSGGAI SSASDIDIQG
 401 NCSAIEFSGN QSLIALGEHI GLTDFVGGGA LAAQGTLTLR NNAVQCVKN
 451 TSKTHGGAIL AGTVIDLNETI SEVAFKQNTA ALTGGALSAN DKVIIANNFG
 501 EILFEQNEVR NHGGAIYCGC RSNPKLEQKD SGENINIIGN SGAITFLKNK
 551 ASVLEVMTQA EDYAGGGALW GHNVLLDSNS GNIQFIGNIG GSTFWIGEYV
 601 GGGAILSTDRT VTISNNSGDV VFKGNGQCL AQKYVAPQET APVESDASST
 651 NKDEKSLNAC SHGDHYPPKT VEEEVPPSSL EEHPVVSSTD IRGGGAILAQ
 701 HIFITDNTGN LRFSGNLGGG EESSTVGDLA IVGGGALLST NEVNVCNQN
 751 VVFSDNVTSN GCDSGGAILA KKVDISANHS VEFVSNSGSK FGGAVCALNE
 801 SVNITDNGSA VSFSKNTRL GGAGVAAPQG SVTICGNQGN IAFKENFVFG

851 SENQRSGGGA IIANSVNIQ DNAGDILFVS NSTGSYGGAI FVGSLVASEG
 901 SNPRTLTITG NSGDILFAKN STQTAASLSE KDSFGGGAIY TQNLKIVKNA
 951 GNVSFYGNRA PSGAGVQIAD GGTVCLEAFG GDILFEGININ FDGSFNIAHL
 1001 CGNDSKIVEL SAVQDKNIIF QDAITYEENT IRGLPDKDVS PLSAPSLIFN
 1051 SKPQDDSAQH HEGTIRFSRG VSKIPQIAAI QEGTLALSQN AELWLAGLKQ
 1101 ETGSSIVLSA GSILRIFDSQ VDSSAPLPTE NKEETLVASG VQINMSSPTP
 1151 NKDKAVDTPV LADIISITVD LSSFVPEQDG TLPLPPEIII PKGTKLHSNA
 1201 IDLKIIDPTN VGYENHALLS SHKDIPLISL KTAEGMTGTP TADASLSNIK
 1251 IDVSLPSITP ATYGHGVWS ESKMEDGRLLV VGWQPTGYKL NPEKQGALVL
 1301 NNWLWSHYTDL RALKQEFAH HTIAQRMELD FSTNVWGSGL GVVEDCQNIG
 1351 EFDGFKHHLT GYALGLDTQL VEDFLIGGCF SQFFGKTESQ SYKAKNDVKS
 1401 YMGAAYAGIL AGPWLIKCAF VYGNINNDLT TDYGTLGIST GSWIGKGFI
 1451 GTSIDYRYIV NPPRFISAIV STVVPFVEAE YVRIDLPEIS EQGKEVRTFQ
 1501 KTRFENVAIP FGFALEHAYS RGSRRAEVNSV QLAYFDVYR KGPVSLITLK
 1551 DAAYSWKSYG VDIPCKAWKA RLSNNTEWNS YLSTYLAFLNY EWREDLIAYD
 1601 FNNGIRIIIF*

A predicted signal peptide is highlighted.

The cp7287 nucleotide sequence <SEQ ID 84> is:

1 ATGGTAGCGA AAAAACAGT ACGATCTTAT AGGTCTTCAT TTTCTCATTC
 20 51 CGTAATAGTA GCAATATTGT CAGCAGGCAT TGCTTTGAA GCACATTCCCT
 101 TACACAGCTC AGAACTAGAT TTAGGTGTAT TCAATAAACAA GTTTGAGGAA
 151 CATTCTGCTC ATGTTGAAGA GGCTCAAACA TCTGTTTAA AGGGATCAGA
 201 TCCGTAAAT CCCCTCTCAGA AAGAATCCGA GAAGGTTTG TACACTCAAG
 25 251 TGCCTCTTAC CCAAGGAAGC TCTGGAGAGA GTTGGATCT CGCCGATGCT
 301 AATTCTTCTAG AGCATTTCAGA GCATCTTTT GAAGAGACTA CAGTATTGG
 351 TATCGATCAA AAGCTGGTT GGTCAGATT AGATACTAGG AATTTCCTCC
 401 AACCCACTCA AGAACCTGAT ACAAGTAATG CTGTAAGTGA GAAAATCTCC
 451 TCAGATACCA AAGAGAATAG AAAAGACCTA GAGACTGAAG ATCCTTCAAA
 501 AAAAAGTGGC CTTAAAGAAG TTCATCAGA TCTCCCTAAA AGTCTGAAA
 551 CTGCAGTAGC AGCTATTCTC AAAGATCTTG AAATCTCAGA AAACATTCA
 601 GCAAGAGATC CTCTTCAGGG TTTAGCATTT TTTTATAAAA ATACATCTTC
 651 TCAGTCTATC TCTGAAAAGG ATTCTTCATT TCAAGGAATT ATCTTTCTG
 701 GTTCAGGAGC TAATTCAAGGG CTAGGTTTG AAAATCTAA GGCGCCGAAA
 751 TCTGGGGCTG CAGTTTATTG TGATCGAGAT ATTGTTTTG AAAATCTTGT
 801 TAAAGGATTG AGTTTTATAT CTITGTGAATC TTTAGAAGAT GGCTCTGCCG
 851 CAGGTGTAAA CATTGTGTG ACCCATTTGTG GTGATGAAAC TCTCACTGAT
 901 TGTGCCACTG GTTGTAGACCT TGAAGCTTTA CGTCTGGTTA AAGATTTTC
 951 TCGTGGAGGA GCTGTTTCA CTGCTCCCAA CCATGAAGTG CAAAATAACC
 1001 TTGCTAGGTGG AATTCTATCC GTTGTAGGCA ATAAAGGAGC TATTGTTGTA
 1051 GAGAAAAATA GTGCTGAGAGA GTCCTAATGGA GGAGCTTTG CTTGCCGAAG
 1101 TTTTGTCTTAC AGTAAACAAACG AAAACACCCG CTTGTGGAAA GAAAATCAAG
 1151 CATTATCAGG AGGAGCCATA TCCTCAGCAA GTGATATTGA TATTCAAGGG
 1201 AACTGTAGCG CTATTGAATT TTCAGGAAAC CAGTCTCTAA TTGCTCTGG
 1251 AGAGCATATA GGGCTTACAG ATTGTAGG TGGAGGAGCT TTAGCTGCTC
 1301 AAGGGACGCT TACCTTAAGA AATAATGCAG TAGTGCAATG TGTAAAAAC
 1351 ACTTCTAAAAA CACATGGTGG AGCTATTCTA GCAGGTACTG TTGATCTCAA
 1401 CGAAACAATT AGCGAAGTTG CCTTTAAGCA GAATACAGCA GCTCTAACTG
 1451 GAGGTGCTTT AAGTGCCTAAT GATAAGGTTA TAATTGCAAA TAACCTTGG
 1501 GAAAATTCTTT TTGAGCAAAA CGAAGTGTGAGG AATCACGGAG GAGCCATT
 1551 TTGTGGATGT CGATCTAATC CTAAGTTAGA ACAAAAGGAT TCTGGAGAGA
 1601 ACATCAATAT TATTGAAAC TCCGGAGCTA TCACTTTTT AAAAATAAG
 1651 GCTTCTGTT TAGAAGTGT GACACAAGCT GAAGATTATG CTGGTGGAGG
 1701 CGCTTTATGG GGGCATAATG TTCTTCTAGA TTCCAATAGT GGGAAATATT
 1751 AATTATAGG AAATATAGGT GGAAGTACCT TCTGGATAGG AGAATATGTC
 1801 GGTGGTGGTG CGATTCTCTC TACTGTAGA GTGACAATT CTAATAACTC
 1851 TGGAGATGTT GTTTTAAAG GAAACAAAGG CCAATGCTTT GCTCAAAAT
 1901 ATGTAGCTCC TCAAGAAACA GCTCCCGTGG AATCAGATGC TTCATCTACA
 1951 AATAAAGACG AGAAGAGCCT TAATGCTTGT AGTCATGGAG ATCATTATCC
 2001 TCCTAAAATC GTAGAAGAGG AAGTGCACC TTCAATTGTTA GAAGAACATC
 2051 CTGTTGTTTC TTCGACAGAT ATTGTGGTG GTGGGGCCAT TCTAGCTCAA
 2101 CATATCTTTA TTACAGATAA TACAGGAAAT CTGAGATTCT CTGGAAACCT
 2151 TGGTGGTGGT GAAGAGTCTT CTACTGTGG TGATTTAGCT ATCGTAGGAG
 2201 GAGGTGCTTT GCTTTCTACT AATGAAGTTA ATGTTTGAG TAACCAAAAT
 2251 GTTGTGTTTT CTGATAACGT GACTTCAAAT GGTTGTGATT CAGGGGGAGC
 2301 TATTTTAGCT AAAAAGTAG ATATCTCCGC GAACCACCTCG GTTGAATTG

	2351	TCTCTAATGG	TTCAGGGAAA	TTCGGTGGTG	CCGTTTGC	TTAAACGAA
5	2401	TCAGTAAACA	TTACGGACAA	TGGCTCGGCA	GTATCATTCT	CTAAAAATAG
	2451	AACACGTCTT	GGCGGTGCTG	GAGTGCAGC	TCCTCAAGGC	TCTGTAACGA
	2501	TTTGTGGAAA	TCAGGGAAAC	ATAGCATT	AAGAGAACTT	TGTTTTG
	2551	TCTGAAAATC	AAAGATCAGG	TGGAGGAGCT	ATCATTGCTA	ACTCTCTGT
	2601	AAATATTCA	GATAACGCAG	GAGATATCCT	ATTTGTAAGT	AACTCTACGG
	2651	GATCTTATGG	AGGTGCTATT	TTTGTAGGAT	CTTGGTTGC	TTCTGAAGGC
10	2701	AGCACACCCAC	GAACGCTTAC	AATTACAGGC	AACAGTGGGG	ATATCCTATT
	2751	TGCTAAAAT	AGCACGCAA	CAGCCGCTTC	TTTATCAGAA	AAAGATTCCT
	2801	TTGGTGGAGG	GGCCATCTAT	ACACAAAACC	TCAAAATTGT	AAAGAATGCA
	2851	GGGAACGTTT	CTTCTATGG	CAACAGAGCT	CCTAGTGGTG	CTGGTGTCCA
	2901	AATTGCAGAC	GGAGGAAC	TTTGTAGA	GGCTTTGGA	GGAGATATCT
	2951	TATTTGAAGG	GAATATCA	TTTGATGGGA	GTTTCAATGC	GATTCACTTA
15	3001	TGCGGAATG	ACTCAAAAT	CGTAGAGCTT	TCTGCTGTT	AAGATAAAA
	3051	TATTATTTTC	CAAGATGCAA	TTACTTATGA	AGAGAACACA	ATTCTG
	3101	TGCCAGATAA	AGATGTCAGT	CCTTTAAGTG	CCCCTTCATT	AATTTTAAC
	3151	TCCAAGGCCAC	AAGATGACAG	CGCTCAACAT	CATGAAGGGA	CGATACGGTT
20	3201	TTCTCGAGGG	GTATCTAAA	TTCCCTCAGAT	TGCTGCTATA	CAAGAGGGAA
	3251	CCTTAGCTTT	ATCACAAAC	GCAGAGCTT	GGTTGGCAGG	ACTTAAACAG
	3301	GAAACAGGAA	GTTCTATCGT	ATTGTC	GGATCTATTC	TCCGTATTT
	3351	TGATTCCCAG	GTTGATAGCA	GTGCCCTCT	TCCTACAGAA	AATAAAGAGG
	3401	AGACTCTTGT	TTCTGCCGGA	GTTCAAATT	ACATGAGCTC	TCCTACACCC
25	3451	AATAAAGATA	AAGCTGTAGA	TACTCCAGTA	CTTGCAGATA	TCATAAGTAT
	3501	TACTGTAGAT	TTGTCTTCAT	TTGTTCC	GCAAGACGGA	ACTCTCCTC
	3551	TTCCTCC	AATTATCATT	CCTAAGGGAA	CAAATTACA	TTCTAATGCC
	3601	ATAGATCTT	AGATTATAGA	TCCTACCA	GTGGGATATG	AAAATCATGC
	3651	TCTTCTAAGT	TCTCATAAAG	ATATTCCATT	AATTCTCTT	AAGACAGCGG
30	3701	AAGGAATGAC	AGGGACCC	ACAGCAGATG	TTCTCTATC	TAATATAAA
	3751	ATAGATGTAT	CTTTACCTTC	GATCACACCA	GCAACGTATG	GTCACACAGG
	3801	AGTTGGTCT	GAAAGTAAA	TGGAAGATGG	AAGACTTGTA	GTCGGTTGGC
	3851	AACCTACGGG	ATATAAGTTA	AATCCTGAGA	AGCAAGGGC	TCTAGTTTG
	3901	AATAATCTCT	GGAGTCATTA	TACAGATCTT	AGAGCTCTA	AGCAGGAGAT
	3951	CTTGCTCAT	CATACGATAG	CTCAAAGAA	GGAGTTAGAT	TTCTCGACAA
35	4001	ATGTCTGGGG	ATCAGGATTA	GGTGTGTTG	AAGATTGTC	GAACATCGGA
	4051	GAGTTTGATG	GGTTCAAACA	TCATCTCACA	GGGTATGCC	TAGGCTTGGA
	4101	TACACAACTA	GTGAGACT	TCTTAATTGG	AGGATGTTTC	TCACAGTCT
	4151	TTGGTAAAAC	TGAAAGCCAA	TCCTACAAAG	CTAAGAACGA	TGTGAAGAGT
	4201	TATATGGGAG	CTGCTTATGC	GGGGATTTA	GCAGGTCC	GGTTAATAAA
40	4251	AGGAGCTTT	GTTPACGGTA	ATATAAACAA	CGATTGACT	ACAGATTACG
	4301	GTACTTTAGG	TATTTCACAA	GGTCATGG	TAGGAAAAGG	GTTTATGCC
	4351	GGCACAAAGCA	TTGATTACCG	CTATATTGTA	AATCTCGAC	GGTTTATATC
	4401	GGCAATCGTA	TCCACAGTGG	TTCCCTTTG	AGAAGCCGAG	TATGTCGTA
	4451	TAGATCTTCC	AGAATTTAGC	GAACAGGGTA	AAGAGGTAG	AACGTTCCAA
45	4501	AAAATCTCGT	TTGAGAATGT	CGCCATTCC	TTTGGATTG	CTTTAGAACAA
	4551	TGCTTATTCCG	CGTGGCTCAC	GTGCTGAAGT	GAACAGTGT	CAGCTTGCTT
	4601	ACGTCTTGT	TGTATATCGT	AAGGGACCTG	TCTCTTGT	TACACTCAAG
	4651	GATGCTGCTT	ATTCTGGAA	GAGTTATGGG	GTAGATATTC	CTTGAAAGC
	4701	TTGGAAGGCT	CGCTTGAGCA	ATAATACGGA	ATGGAATTCA	TATTTAAGTA
	4751	CGTATTAGC	GTTTAATTAT	GAATGGAGAG	AAGATCTGAT	AGCTTATGAC
50	4801	TTCAATGGTG	GTATCCGTAT	TATTTCTAG		

The PSORT algorithm predicts an inner membrane location (0.106).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 42A.

The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 42B) and for FACS analysis (Figure 42C). A his-tagged protein was also expressed.

55 The cp7287 protein was also identified in the 2D-PAGE experiment and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7287 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 43

The following *C.pneumoniae* protein (PID 4377105) was expressed <SEQ ID 85; cp7105>:

```

 1  MSLYQKWWNS QLKSLCYST VAALIFMIPS QESFADSLID LNLGLDPSVE
 51  CLSGDGAFSV GYFTKAGSTP VEYQPFKYDV SKKTFILSV ETANQSGYAY
 101  GISYDGTITV GTCISLGAKY NGAKWSADGT LTPLTGITGG TSHTEARAIS
 151  KDTQVIEGFS YDASGQPKAV QWASGATTVT QLADISGGSR SSYAYAISDD
 201  GTIIVGSMES TITRKTTAVK WVNNVPTYLG TLGGDASTGL YISGDGTVIV
 251  GAANTATVTN GNQESHAYMY KDNQMKD*

```

The cp7105 nucleotide sequence <SEQ ID 86> is:

```

10      1  GTGAGTCTAT ATCAAAAATG GTGGAACAGT CAGTTAAAGA AGAGCCTCTG
      51  CTATTCGACT GTTGCTGCTC TAATATTTAT GATTCCTTCT CAAGAATCCT
      101  TTGCAGATAG TCTTATAGAT TTAATTTAG GTTTAGATCC TTCGGTCGAA
      151  TGTCTGTCAG GAGATGGTGC ATTTCCTGTT GGGTATTTA CTAAGGGCGGG
      201  ATCGACTCCC GTAGAACATC AGCCGTTAA ATACGACGTA TCTAAGAAGA
      251  CATTCAACAAT CCTTTCCGT ACAAACGGCAA ATCAGAGCGG CTATGCTTAC
      301  GGAATCTCCT ACGATGGCAC GATCACTGTA GGAACGTGTA GCCTAGGTGC
      351  AGGAAAATAT AACGGGCCAA AATGGAGTGC GGATGGCACT TTAACACCCT
      401  TAACTGGAAT CACGGGGGG ACGTACACATA CGGAAGCGCG TGCGATTCT
      451  AAGGATACTC AGGTGATCGA GGGTTCTCA TATGATGCTT CAGGGCAACC
      501  CAAGGCTGTG CAGTGGCAA GCGGAGCGAC TACAGTAACA CAATTAGCAG
      551  ATATTCAGG AGGCTCTAA AGCTCTTATG CGTATGCTAT ATCTGATGAT
      601  GGCACGATTA TTGTTGGTC TATGGAGAGC ACGATAACAA GGAAAACATC
      651  AGCTGTAAAA TGGGTAATA ATGTTCTAC GTATCTGGGA ACCTTAGGAG
      701  GAGATGCTTC TACAGGTCTT TATAATTCCTG GAGACGGCAC CGTGATTGTA
      751  GGTGCGGCAA ATACAGCAAC TGTAACCAAT GGGAAATCAGG AATCCACCGC
      801  CTATATGTAT AAAGATAACC AAATGAAAGA TTGA

```

The PSORT algorithm predicts an inner membrane location (0.100).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 43A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 43B) and for FACS analysis (Figure 43C). A his-tagged protein was also expressed.

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7105 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 44

The following *C.pneumoniae* protein (PID 4376802) was expressed <SEQ ID 87; cp6802>:

```

 1  MSNQLQPCIS LGCVSYINSF PLSQLIKRN DIRCVLAPPA DLLNLILLIEGK
 51  LDVALTSSLG AISHNLGYVP GFGIAANQRI LSVNLYAAPT FFNSPQPRIA
 101  ATLESRSSIG LLKVLCRHLW RIPTPHILRF ITTKVLRQTP ENYDGLLLIG
 151  DAALQHPVLP GFVTYDLASG WYDLTKLPFV FALLLHSTSW KEHPLPNLAM
 201  EEAQQFESS PEEVLKEAHQ HTGLPPSLLQ EYYALCQYRL GEEHYESFEK
 251  FREYYGTLQ* QARL*

```

A predicted signal peptide is highlighted.

The cp6802 nucleotide sequence <SEQ ID 88> is:

```

45      1  ATGTCTAACCC AACTCCAGCC ATGTATAAGC TTAGGCTGGG TAAGTTATAT
      51  TAATTCCTTT CCGCTGTCCC TACAACTCAT AAAAAGAAAC GATATTCGCT
      101  GTGTTCTTGC TCCCCCTGCA GACCTCCTCA ACTTGCTAAT CGAAGGGAAA
      151  CTCGATGTTG CTTTGACCTC ATCCCTAGGA GCTATCTCTC ATAACCTTGGG
      201  GTATGTCCCC GGCTTTGGAA TTGAGCAAA CCAACGTATC CTCAGTGTAA

```

251 ACCTCTATGC AGCTCCCACT TTCTTTAACT CACCGCAACC TCGGATTGCC
 301 GCAAACTTAG AAAGTCGCTC CTCTATAGGA CTCTTAAAG TGCTTTGTCG
 351 TCATCTCTGG CGCATCCAA CCTCTCATAT CCTAAGATTC ATAACATCAA
 401 AAGTACTCAG ACAAAACCCCT GAAAATTATG ATGGCCTCCT CCTAATCGGA
 451 GATGCAGCGC TACAACATCC TGTACTTCCT GGATTTGTAA CCTATGACCT
 501 TGCCTCGGGG TGGTATGATC TTACAAAGCT ACCTTTTGTA TTTGCTCTTC
 551 TTCTACACAG CACCTCTGG AAAGAACATC CCCTACCCAA CCTTGCATG
 601 GAAGAAGGCC TCCAACAGTT CGAACATTC CCCGAAGAAG TCCTTAAAGA
 651 AGCTCATCAA CATACAGGTC TGCCCCCTTC TCTTCCTCAA GAATACTATG
 701 CCCTATGCCA GTACCGTCTA GGAGAAGAAC ACTACGAAAG CTTTGAAAAAA
 751 TTCCGGGAAT ATTATGAAAC CCTCTACCAA CAAGCCGAC TGTAA

The PSORT algorithm predicts an inner membrane location (0.060).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 44A.

The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 44B) and for FACS analysis (Figure 44C). A his-tagged protein was also expressed.

These experiments show that cp6802 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 45

The following *C.pneumoniae* protein (PID 4376390) was expressed <SEQ ID 89; cp6390>:

20 1 **MVFSYYCMGL FFFSGAISSC GLLVSLGVGL GLSVLGVLLL LLAGLLLLFKI**
 51 **QSMI**REVPKA PDLLDLEDAS ERLRVKASRS LASLPKEISQ LESYIRSAAN
 101 DLNTIKTWPH KDQRLVETVS RKLERLAAAQ NYMISELCEI SEILEEEEHH
 151 LILAQESLEW IGKSLFSTFL DMESFLNLSH LSEVRPYLAV NDPRLLEITE
 201 ESWEVVSHFI NVTSAFKKAQ ILFKNNEHSR MKKKLESVQE LLETFIYKSL
 251 KRSYRELGCL SEKMRIIHND PLFPWVQDQQ KYAHAKNEFG EIARCLEEFE
 301 KTFFWLDEEC AISYMDCWDF LNESIQNKKS RVDRDYISTK KIALKDRART
 351 YAKVLLLEENP TTEGKIDLQD AQRAYERQSQ EFYTLEHTET KVRLEALQOC
 401 FSDLREATNV RQVRFTNSEN ANDLKESFEK IDKERVRYQK EQRLYWETID
 451 RNEQELREEI GESLRLQNRR KGAGRAGYDAG RLKGLLRQWK KNLRDVEAHL
 501 EDATMDFEHE VSKSELCSVRL ARLEVLEEE MDMSPKVADI EELLSYEERC
 551 ILPIRENLER AYLQYNKCSE ILSKAKFFF EDEQLLVSEA NLREVGAQLK
 601 QVQGKCQERA QKFAIFEKHI QEQLIKEQ VRSFDLAGVG FLKSELLSIA
 651 CNLYIKAVVK ESIPVDVPCM QLYYYSYEDN EAVVRNRLLN MTERYQNFKR
 701 SINSIQFNND VLLRDPVYQP EGHETRLKER ELQETTLCK KLKVAQDRLS
 751 ELESRLSRR

A predicted signal peptide is highlighted.

The cp6390 nucleotide sequence <SEQ ID 90> is:

40 1 TTGGTATTCT CATACTATTG CATGGGATTA TTTTTTTCT CTGGAGCTAT
 51 TTCTAGTTGT GGTCTTTAG TGTCTCTAGG AGTTGGTTA GGACTTAGTG
 101 TTTAGGAGT ACTTTTACTT CTCTTAGCAG GTCTTTGCT TTTAAGATC
 151 CAAAGTATGC TTGAGAGGT GCCTAAGGCT CCTGATCTAT TAGATTTAGA
 201 AGATGCAAGT GAACGGCTTA GAGTAAAGGC TAGCCGTTCT TTAGCAAGCC
 251 TCCCAGAGGA AATCAGTCAG CTAGAGAGCT ACATTCGTC TGCAAGCTAAT
 301 GATCTAAATA CAATTAAGAC TTGGCCGCAT AAAGATCAA GACTCGTCGA
 351 GACCCTGTCA CGAAAATTAG AGCGTCTGGC AGCTGCTCAA AACTATATGA
 401 TTCTGAACT CTGCGAGATT AGTGAGATTG TTGAGGAAGA GGAGCATCAT
 451 CTAATTTGG CTCAGGAATC TCTAGAATGG ATAGGTAAGA GTCTATTTTC
 501 TACCTTTCTG GACATGGAACT CTTTTTTAAA TTGAGCCAT CTATCTGAAG
 551 TGGTCCCGTA CTTAGCTGTA AATGATCCTA GATTATTAGA AATTACCGAA
 601 GAATCTTGGG AAGTAGTGTAG TCATTTCTATA AATGTAACGT CTGCTTTAA
 651 GAAAGCTCAG ATTCTTTTA AGAACACAACGA ACATTCTCGG ATGAAAGAAGA
 701 AGTTAGAAAG TGTTCAAGAG TTACTGGAAA CATTATTTA TAAGAGTTTA
 751 AAGAGAAGTT ATCGAGAATT AGGATGCTTA AGTGAAAAGA TGAGAATCAT
 801 TCACGACAAT CCTCTCTTCC CTTGGGTGCA AGATCAGCAG AAGTATGCTC
 851 ATGCTAAGAA TGAATTGGG GAGATTGCGC GGTGTTAGA GGAGTTGAA
 901 AAGACGTTCT TCTGGTTGGA TGAGGAGTGT GCTATTTCTT ACATGGACTG

951 TTGGGATTTT CTAAATGAGT CTATTCAGAA TAAGAAGTCC AGAGTAGATC
 1001 GAGATTATAT ATCCACGAAG AAAATTGCAT TAAAGGATAG AGCCCGCACT
 1051 TATGCTAAGG TTCTTTAGA AGAGAACCG ACTACAGAGG GTAAAATAGA
 1101 TTTGCAAGAC GCTCAAAGAG CCTTGAGCG TCAAAGTCAG GAGTTTATA
 1151 CACTAGAGCA TACGGAACAA AAGGTGAGAC TAGAAGCACT TCAACAGTGC
 1201 TTCTCGGATC TTAGGGAGGC GACGAACGTA AGGCAAGTTA GGTTTACAA
 1251 TTCTGAAAAT GCGAATGATT TAAAGGAGAG TTTCGAGAAG ATAGATAAAAG
 1301 AGCGTGTGCG ATATCAAAAA GAGCAAAGGC TCTATTGGGA AACAAATAGAT
 1351 CGCAATGAGC AAGAGCTAG GGAAGAGATT GGGGAGTCGC TTCGTTTACA
 1401 AAATCGGAGA AAAGGGTATA GGGCTGGATA TGATGCTGGG CGTTTAAAG
 1451 GTTGTGCG TCAGTGGAA AAAATCTCC GCGATGTGGA AGCCCACCTT
 1501 GAAGATGCAA CTATGGATT TGACCATGAA GTAAGCAAGA GCGAATTGTG
 1551 CAGTGTGCG GCGAGGCTCG AGGTCTAGA AGAAGAGCTG ATGGATATGT
 1601 CTCCTAAAGT TGCGGATATA GAAGAGTTGT TGTCTATGA AGAGCGTTGT
 1651 ATTCTTCTA TTAGGGAAA TTAGAAAAGG GCATACCTCC AATATAATAA
 1701 GTGTTCTGAA ATTTTATCCA AGGCAAAGTT CTTCTTCCG GAAGACGAGC
 1751 AATTGCTAGT TTCGGAAAGCG AATCTAAGAG AGGTGGGTGC CCAGTTAAA
 1801 CAAGTACAGG GAAAATGTCAGAGAGGAGGCC CAAAAGTTCG CAATATTGAG
 1851 AAAGCATATT CAGGAGCAGA AAAGCCTTAT TAAAGAGCAA GTGCCGAGTT
 1901 TTGATCTAGC GGGAGTTGGG TTTTAAAGA GTGAGCTTCT TAGTATTGCT
 1951 TGTAAACCTT ATATAAAAGGC GGTGTTAAAG GAGTCTATAC CAGTTGATGT
 2001 GCCTTGTATG CAGTTATATT ATAGTTATTA CGAAGATAAT GAAGCTGTAG
 2051 TGCGAAACCG CCTTTAAAT ATGACGGAGA GGTATCAAAA TTTTAAAGG
 2101 AGTTTGAATT CCATACAATT TAATGGTGCAG GTTCTTTAC GGGATCCGGT
 2151 CTATCAACCT GAAGGTCATG AGACCAGGCT AAAGGAACGG GAGCTACAAG
 2201 AAACAACCTT GTCTTGTAAAG AAATTAAAAG TGGCTCAAGA TCGTCTTCT
 2251 GAATTAGAGT CAAGGCTGTC TAGGAGATAG

The PSORT algorithm predicts a periplasmic location (0.932).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 45A.

30 The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 45B) and for FACS analysis (Figure 45C). A his-tagged protein was also expressed.

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6390 is a surface-exposed and immunoaccessible protein, and that it 35 is a useful immunogen. These properties are not evident from the sequence alone.

Example 46

The following *C.pneumoniae* protein (PID 4376272) was expressed <SEQ ID 91; cp6272>:

1 MKRCFLFLAS FVLMGSSADA LTHQEAVKKK NSYLSHFKSV SGIVTIEDGV
 51 LNIHNNLRIQ ANKVYVENTV GQSLKLVAHG NVMVNRYAKT LVCDYLEYYE
 101 DTDSCLLTNG RFAMYPWFLG GSMITLTPET IVIRKGYIST SEGPKKDLCL
 151 SGDYLEYSSD SLLSIGKTTL RVCRIPIFL PPFSIMPMEI PKPPINFRGG
 201 TGGFLGSYLG MSYSPISRKH FSSTFFLDSF FKHGVMGMFN LHCSQKQVPE
 251 NVFNMKSYA HRLAIDMAEA HDRYRLHGDF CFTHKHVNFS GEYHILSDSWE
 301 TVADIFPNNF MLKNTGPTRV DCTWNNDNYFE GYLTSVKVN SFQANQELP
 351 YLTLRQYPIS IYNTGVYLEN IVECGYLNFA FSDHIVGENF SSLRLAARPK
 401 LHKTVPPLIG TLSSTLGSSL IYYSDVPEIS SRHSQLSAKL QLDYRFLLHK
 451 SYIQRRHIIE PFVTFITETR PLAKNEDHYI FSIQDAFHSL NLLKAGIDTS
 501 VLSKTNPRFP RIHAKLWPTH ILSMNTESKPT FPKTACELSL PFGKKNTVSL
 551 DAEWIWKKHC WDHMNIRWEW IGNNDNVAMTL ESLHRSKYSI IKCDRENFIL
 601 DVSRPIDQLL DSPLSDHRNL ILGKLFVRPH PCWNYRLSLR YGWHRQDTPN
 651 YLEYQMILGT KIFEHWQLYG VYERREADSR FFFFILKLDKP KKPPF*

A predicted signal peptide is highlighted.

The cp6272 nucleotide sequence <SEQ ID 92> is:

1 ATGAAACGTT GCTTCTTATT TCTAGCTTCC TTTGTTCTTA TGGGTTCCCTC

5 51 AGCTGATGCT TTGACTCATC AAGAGGCTGT GAAAAAGAAA AACTCCTATC
 101 101 TTAGTCACTT TAAGAGTGT TCTGGGATTG TGACCATCGA AGATGGGTA
 151 151 TTGAATATCC ATAACAACCT GCGGATACAA GCCAATAAG TGTATGTAGA
 201 201 AAATACTGTG GGTCAAAAGCC TGAAGCTTGT CGCACATGGC AATGTTATGG
 251 251 TGAACATAG GGCAAAAACC CTAGTTGTG ATTACCTAGA GTATTACGAA
 301 301 GATACAGACT CTTGTCTTCT TACTAATGGA AGATTCGCGA TGTATCCTTG
 351 351 GTTCTAGGG GGGTCTATGA TCACCTAAC CCCAGAAACC ATAGTCATTIC
 401 401 GGAAGGGATA TATCTCTACC TCCGAGGGTC CCAAAAAAGA CCTGTGCCTC
 451 451 TCCGGAGATT ACCTGGAATA TTCTTCAGAT AGTCTCTTT CTATAGGGAA
 501 501 GACAACATTA AGGGTGTGTC GCATTCGAT ACTTTCTTA CCTCCATTTT
 551 551 CTATCATGCC TATGGAGATC CCTAAGCCTC CGATAAAACTT TCGAGGAGGA
 601 601 ACAGGAGGAT TTCTGGGATC CTATTGGGG ATGAGCTACT CGCCGATTTC
 651 651 TAGGAAGCAT TTCTCCTCGA CATTTCTT GGATAGCTTT TTCAAGCATG
 701 701 GCGTCGGCAT GGGATTCAAC CTCCATTGTG CCTCGAAGCA GGTTCTGAG
 751 751 AATGCTTCA ATATGAAAAG CTATTATGCC CACCGCCTTG CTATCGATAT
 801 801 GGCAGAAGCT CATGATCGCT ATCGCCTACA CGGAGATTTC TGCTTCACGC
 851 851 ATAAGCATGT AAATTTTCT GGAGAATACC ATCTCAGCGA TAGTTGGAA
 901 901 ACTGTTGCTG ACATTTCCC CAACAACCTC ATGTTGAAA ATACAGGCC
 951 951 CACACGTGTC GATTGCACTT GGAATGACAA CTATTTGAA GGTTATCTCA
 1001 1001 CCTCTTCTGT TAAGGTAAC TCTTCCAAA ATGCCAACCA AGAGCTCCCT
 1051 1051 TATTTAACAT TAAGGCAGTA CCCGATTTC ATTTATAATA CGGGAGTGTA
 1101 1101 CCTTGAAAAC ATCGTAGAAT GTGGGTATTT AACTTTGCT TTTAGCGATC
 1151 1151 ATATCGTTGG CGAGAATTTC TCTTCACTAC GTCTGCTGC GCGCCCTAAG
 1201 1201 CTCCATAAAAA CTGTGCTCT ACCTATAGGA ACGCTCTCT CCACCCCTAGG
 1251 1251 GAGTTCTCTG ATTTACTATA GCGATGTTCC TGAGATCTCC TCGGCCATA
 1301 1301 GTCAGCTTTC CGCGAAGCTA CAACTGATT ATCGCTTTCT ATTACATAAG
 1351 1351 TCCTACATTC AAAGACGCCA TATTATAGAG CGGTTCGTTA CTTTCATTAC
 1401 1401 AGAGACTCGT CCTCTAGCTA AGAATGAAGA TCATTATATC TTTTCTATTTC
 1451 1451 AAGATGCCTT TCACTCCTTA AACCTCTGA AAGCGGGTAT AGATAACCTCG
 1501 1501 GTACTGAGTA AGACTAACCC TCGATTCCCG AGAATCCATG CGAAGCTGTG
 1551 1551 GACTACCCAC ATCTTGAGCA ATACAGAAAG CAAACCCACG TTTCCAAAAA
 1601 1601 CTGCATGCGA GCTATCTCTA CCTTTGGAA AGAAAATAC AGTCTCCTTA
 1651 1651 GATGCTGAAT GGATTTGGAA AAAGCACTGT TGGGATCACA TGAACATACG
 1701 1701 TTGGGAGTGG ATCGGAAATG ACAATGTGGC TATGACTCTA GAATCCCTGC
 1751 1751 ATAGAAGCAA ATACAGCCTG ATTAAGTGTG ACAGGGAGAA CTTCATTTTA
 1801 1801 GATGTCAGCC GTCCCATTGA CCAGCTTTA GACTCCCCCTC TCTCTGATCA
 1851 1851 TAGGAATCTC ATTTTAGGGA AATTATTGT ACGACCTCAT CCCTGTGTTGGA
 1901 1901 ATTACCGCTT ATCCTTACGC TATGGCTGGC ATCGCCAGGA CACTCCGAAC
 1951 1951 TACCTAGAAT ACCAGATGAT TCTAGGGACG AAGATCTTCG AACATTGGCA
 2001 2001 GCTCTATGGG GTGTATGAAC GCCGAGAAGC AGATAGTCGA TTTTTCTTCT
 2051 2051 TCTTAAAGCT CGACAAACCT AAAAAACCTC CCTTCTAA

The PSORT algorithm predicts an outer membrane location (0.48).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 46A.

The recombinant protein was used to immunise mice, whose sera were used in a Western blot and for

FACS analysis (Figure 46B). A his-tagged protein was also expressed.

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6272 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

50 Example 47

The following *C.pneumoniae* protein (PID 4377111) was expressed <SEQ ID 93; cp7111>:

5 1 MFEAVIADIQ AREILDERSGY PTLHVKVTTG TGSGVGEARVP SGASTGKKEA
 55 51 LEFRDTDSPR YQGKGVLQAV KNVKEILFPL VKGCSVYEQS LIDSLMMDS
 101 101 GSPNKETLGA NAILGVSLAT AHAAAATLRR PLYRYLGGCF ACSLPCPMNN
 151 151 LINGGMHAQN GLEFQEFLR PIGASSIKEA VNMGADVFHT LKKLLHERGL
 201 201 STGVCDEGGF APNLASNEEA LELLLAIKEK AGFTPGKDLS LALDCAASSF

251 YNVKTGTYDG RHYEEQIAIL SNLCDRYPID SIEDGLAED YDGWALLTEV
 301 LGEKVQIVGD DLFVTNPRLI LEGISNGLAN SVLIKPNQIG TLTETVYAIK
 351 LAQMAGYTTI ISHRSGETTD TTIADLAVAF NAGQIKTGSL SRSERVAKYN
 401 RLMEIEELG SEAIFTDSNV FSYEDSEE*

5 A predicted signal peptide is highlighted.

The cp7111 nucleotide sequence <SEQ ID 94> is:

1 ATGTTTGAAG CTGTCATTGC CGATATCCAG GCTAGGGAAA TCTTGGATTG
 51 TCGCGGGTAT CCCACTTAC ATGTTAAAGT AACCACTAGC ACAGGTTCTG
 101 TTGGAGAACG TCAGGTTCCCT TCAGGAGCAT CCACAGGGAA AAAAGAAGCC
 151 TTAGAGTTTC GTGATACAGA TTCTCCTCGT TATCAAGGCA AAGGGGTTTT
 201 GCAAGCTGTAA AAAAACGTAA AAGAAATTCT TTTTCCCCCTC GTCAAGGGAT
 251 GTAGTGTATA TGAGCAATCC TTAATTGATT CTCTGATGAT GGATTCTGAC
 301 GGCTCTCCGA ACAAAAGAAAC TCTAGGGGCC AATGCTATT TAGGAGTCTC
 351 TCTAGCTACA GCACATGCAG CAGCAGCAAC ACTACGCAGA CCTCTGTATC
 401 GTTATTAGG AGGGTGTGTT GCCTGCAGTC TTCCCTGTCC TATGATGAAT
 451 CTGATCAATG GAGGCATGCA TGCCGATAAC GGCTTGGAGT TCCAAGAATT
 501 TATGATCCGT CCTATTGGAG CCTCTTCCAT CAAAGAAGCT GTCAACATGG
 551 GTGCTGACGT TTTTCATACT TTGAAAAAAAT TACTCCATGA AAGAGGCTTA
 601 TCTACTGGAG TGGGTGACGA AGGAGGCTTC GCCCGAATC TTGCTTCTAA
 651 TGAAGAAGCT CTAGAGCTCC TATTGCTGGC TATTGAAAAA GCAGGCTTTA
 701 CTCCAGGAAA AGATATATCG CTAGCCTTAG ACTGCGCAGC ATCCTCATTC
 751 TATAACGTAA AAACAGGCAC GTATGATGGG AGGCACTATG AAGAGCAAAT
 801 CGCAATCCTT TCTAATTAT GTGATCGCTA TCCTATAGAC TCCATAGAAG
 851 ATGGTCTTGC TGAAGAAGAC TATGACGGGT GGGCCTTGTGTT AACTGAAGTT
 901 CTGGAGAGAA AAGTACAGAT TGTGGGTGAT GACCTATTG TTACAATCC
 951 GGAATTAAATA TTAGAGGGTA TTAGCAATGG ATTAGCGAAC TCTGTGTTGA
 1001 TAAACACAAA TCAGATAGGG ACCGTTACTG AAACAGTGTAA TGCTATCAAG
 1051 CTGCGCAAA TGGCTGGCTA TACTACAATT ATTTCATC GCTCAGGAGA
 1101 AACTACGGAC ACTACGATTG CAGATCTTGC TGTTGCCTTC AACGCCGGTC
 1151 AAATCAAAAC AGGCTCTTA TCACGTTCTG AGCGTGTGTC AAAATACAAT
 1201 AGACTCATGG AAATTGAAGA AGAGCTTGGAA TCCGAAGCAA TTTTCACAGA
 1251 TTCTAATGTA TTTCTTAC GAGGATTCT GAGGAATAG

The PSORT algorithm predicts an inner membrane location (0.100).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 47A.

35 The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 47B) and for FACS analysis (Figure 47C). A his-tagged protein was also expressed.

The cp7111 protein was also identified in the 2D-PAGE experiment and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7111 is a surface-exposed and immunoaccessible protein, and that it 40 is a useful immunogen. These properties are not evident from the sequence alone.

Example 48

The following *C.pneumoniae* protein (PID 4455886) was expressed <SEQ ID 95; cp0010>:

1 MKSQFSWLVL SSTLACFTSC STVFAATAEN IGPDSFDGS TNTGTYTPKN
 51 TTTGIDYTLT GDITLQLNLD SAALTKGCF DTTELSFAG KGYSLFLNI
 101 KSSAEGAALS VTTDKNLSLT GFSSLTFLAA PSSVITTPSG KGAVKCGDL
 151 TFDNNNGTILF KQDYCEENG AISTKNLSLK NSTGSISFEG NKSSATGKKG
 201 GAICATGTVD ITNNNTAPTLF SNNIAEAAGG AINSTGNCTI TGNTSLVFSE
 251 NSVTATAGNG GALSGDADVT ISGNQSVTF GSQAVANGGA IYAKKLTLAS
 301 GGGGVSPFLT IIVQGTTAGN GGAISILAAG ECLSLAEAGD ITFNGNAIVA
 351 TPPQTTRKNS IDIGSTAKIT NLRAISGHSI FFYDPITANT AADSTDTLNL
 401 NKADAGNSTD YSGSIVFSGE KLSDEDEAKVA DNLTSTLKQP VTLTAGNLVL
 451 KRGVTLDTKG FTQTAGSSVI MDAGTTLKAS TEEVTLTGLS IPVDSLGEVK
 501 KVIAASAAS KNVALSGPIL LLDNQGNAYE NHDLGKTQDF SFVQLSALGT

551 ATTTDVPAVP TVATPTHYGY QGTWGMTWVD DTASTPKTKT ATLAWTNTGY
 601 LPNPERQGPL VPNSLWGGSFS DIQAIQGVIE RSALTLCSDR GFWAAGVANF
 651 LDKDKKGEKR KYRHKSGGYA IGGAAQTCSE NLISFAFCQL FGSDKDFLVA
 701 KNHTDTYAGA FYIQHITECS GFIGCLLDKL PGSWSHKPLV LEGQLAYSHV
 751 SNDLKTKYTA YPEVKGSWGN NAFNMMLGAS SHSYPEYLHC FDTYAPYIKL
 801 NLTYIRQDSF SEKGTEGRSRF DDSNLFNLSL PIGVKFEKFS DCNDFSYDLT
 851 LSYVPDLIRN DPKCTIALVI SGASWETYAN NLARQALQVR AGSHYAFSPM
 901 FEVLGQFVFE VRGSSRIYNV DLGGKFQF*

A predicted signal peptide is highlighted.

10 The cp0010 nucleotide sequence <SEQ ID 96> is:

	1	ATGAAATCGC AATTTCCTG GTTAGTGCTC TCTTCGACAT TGGCATGTTT
	51	TACTAGTTGT TCCACTGTTT TTGCTGCAAC TGCTGAAAAT ATAGGCCCT
15	101	CTGATAGCTT TGACGGAAGT ACTAACACAG GCACCTATAC TCCTAAAAAT
	151	ACGACTACTG GAATAGACTA TACTCTGACA GGAGATATAA CTCTGCAAAA
	201	CCTTGGGGAT TCGGCAGCTT TAACGAAGGG TTGTTTTCT GACACTACGG
	251	AATCTTTAAG CTTTGGCCGGT AAGGGGTACT CACTTCTT TTTAAATATT
	301	AAGTCTAGTG CTGAAGGCCG AGCACTTTCT GTTACAACCTG ATAAAAATCT
	351	GTCGCTAACAA GGATTTTCGA GTCTTACTTT CTTAGCGGGC CCATCATCGG
20	401	TAATCACAAAC CCCCTCAGGA AAAGGTGCAG TTAAATGTGG AGGGGATCTT
	451	ACATTTGATA ACAATGGAAC TATTTTATTAA AAACAAGATT ACTGTGAGGA
	501	AAATGGCGGA GCCATTCTA CCAAGAATCT TTCTTGAAG AACAGCACGG
	551	GATCGATTTTC TTTTGAAGGG AATAAATCGA GCGCAACAGG GAAAAAAAGGT
	601	GGGGCTATTT GTGCTACTGG TACTGTAGAT ATTACAAATA ATACGGCTCC
25	651	TACCCCTCTTC TCGAACAAATA TTGCTGAAGC TGCAAGGTGGA GCTATAAATA
	701	GCACAGGAAA CTGTACAATT ACAGGGAAATA CGTCTCTTGT ATTTCTGAA
	751	AATAGTGTGA CAGCGACCAGC AGGAAATGGA GGAGCTCTT CTGGAGATGC
	801	CGATGTTTACCA ATATCTGGGA ATCAGAGTGT AACTTTCTCA GGAAACCAAG
	851	CTGTAGCTAA TGGCGGGAGCC ATTATATGCTA AGAAGCTTAC ACTGGCTTCC
30	901	GGGGGGGGGGG CGGTATCTCC TTTCTAAACA ATAaTAGTCC AAGGTACCC
	951	TGCAGGTAAT GGTGGAGCCA TTTCTATACT GGCAGCTGGA GAGTGTAGTC
	1001	TTTCAGCAGA AGCAGGGGAC ATTACCTTCA ATGGGAATGC CATTGTTGCA
	1051	ACTACACCAC AAACCTACAA AAGAAATTCT ATTGACATAG GATCTACTGC
	1101	AAAGATCACG AATTTACGTG CAATATCTGG GCATAGCATC TTTTCTACG
35	1151	ATCCGATTAC TGCTAACACG GCTGCGGATT CTACAGATAC TTTAAATCTC
	1201	AATAAGGCTG ATGCAGGTA TAGTACAGAT TATAGTGGGT CGATTGTTTT
	1251	TTCCTGGTAA AAGCTCTCTG AAGATGAAGC AAAAGTTGCA GACAACCTCA
	1301	CTCTACGCT GAAGCAGCCT GTAACCTAA TGCAGGAAA TTTAGTACTT
	1351	AAACGTGGTG TCACTCTCGA TACGAAAGGC TTACTCAGA CCGCGGGTTC
40	1401	CTCTGTTATT ATGGATGCGG GCACAACGTT AAAAGCAAGT ACAGAGGAGG
	1451	TCACTTTAAC AGGTCTTCC ATTCTCTGTAG ACTCTTTAGG CGAGGGTAAG
	1501	AAAGTTGTAA TTGCTGCTTC TGCAAGCAAGT AAAATGTAG CCCTTGTG
	1551	TCCGATTCTT CTTTGGATA ACCAAGGGAA TGCTTATGAA AATCACGACT
	1601	TAGGAAAAAAC TCAAGACTTT TCATTTGTGC AGCTCTCTGC TCTGGGTACT
45	1651	GCAACAACTA CAGATGTTCC AGCGGTTCT ACAGTAGCAA CTCCTACGCA
	1701	CTATGGGTAT CAAGGTACTT GGGGAATGAC TTGGGTTGAT GATACCGCAA
	1751	GCACTCCAAA GACTAACACA GCGACATTAG CTTGGACCAA TACAGGCTAC
	1801	CTTCCGAATC CTGAGCGTCA AGGACCTTA GTTCCAATAA GCCTTGGGG
	1851	ATCTTTTCA GACATCCAAG CGATTCAAGG TGTCAATAGAG AGAAGTGCTT
50	1901	TGACTCTTTG TTCAGATCGA GGCTTCTGGG CTGCGGGAGT CGCCAATTTC
	1951	TTAGATAAAAG ATAAGAAAGG GGAAAAACGC AAATACCGTC ATAAATCTGG
	2001	TGGATATGCT ATCGGAGGTG CAGCGCAAAC TTGTTCTGAA AACTTAATTA
	2051	GCTTTGCCTT TTGCCAACTC TTTGGTAGCG ATAAAGATTT CTTAGTCGCT
	2101	AAAAATCATA CTGATACCTA TGCAAGGAGCC TTCTATATCC AACACATTAC
55	2151	AGAATGTAGT GGGTCATAG TTGTCCTCT AGATAAAACTT CCTGGCTCTT
	2201	GGAGTCATAA ACCCTCGTT TTAGAAGGGC AGCTCGCTTA TAGCCACGTC
	2251	AGTAATGATC TGAAGACAAA GTATACTGCG TATCCTGAGG TGAAAGGTT
	2301	TTGGGGGAAT AATGCTTTA ACATGATGTT GGGAGCTCT TCTCATTCTT
	2351	ATCCTGAATA CCTGCATTGT TTGATACCT ATGCTCCATA CATCAAAC
60	2401	AATCTGACCT ATATAACCTCA GGACAGCTTC TCGGAGAAAG GTACACAGG
	2451	AAGATCTTTT GATGACAGCA ACCTCTCAA TTATCTTGT CCTATAGGGG
	2501	TGAAGTTTGA GAAGTTCTCT GATTGTAATG ACTTTCTTA TGATCTGACT
	2551	TTATCCTATG TTCTGTATCT TATCCGCAAT GATCCCCAAT GCACTACAGC
	2601	ACTTGTAAATC AGCGGAGCCT CTTGGGAAAC TTATGCCAAAT AACTTAGCAC
	2651	GACAGGCCCTT GCAAGTGCCT GCAGGCAGTC ACTACGCCCTT CTCTCCTATG
65	2701	TTTGAAGTGC TCGGCCAGTT TGTCTTTGAA GTTCGTGGAT CCTCACGGAT

2751 TTATAATGTA GATCTTGGGG GTAAGTTCCA ATTCTAG

The PSORT algorithm predicts an outer membrane location (0.922).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 48A.

The recombinant protein was used to immunise mice, whose sera were used in a Western blot

5 (Figure 48B) and for FACS analysis (Figure 48C). A his-tagged protein was also expressed.

The cp0010 protein was also identified in the 2D-PAGE experiment and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp0010 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

10 Example 49

The following *C.pneumoniae* protein (PID 4376296) was expressed <SEQ ID 97; cp6296>:

1	MEEVSEYLQQ VENQLESCSK RLTKMETFAL GVRLEAKEEI ESIILSDVVN
51	RFEVLCRDIE DMLSRVEEIE RMLRMAELPL LPIKEALTKA FVQHNSCKEK
101	LTKVEPYFKE SPAYLTSEER LQSLNQTLQR AYKESQKVSG LESEVRACRE
151	QLKDQVRQFE TQGVSLIKEE ILFVTSTFRT KFSYHSFRLH VPCMRLYEEY
201	YDDIDLERTR ARWMAMSERY RDAFQAFQEM LKEGLVEEAQ ALRETEYWLY
251	REERKSKKH*

The cp6296 nucleotide sequence <SEQ ID 98> is:

1	ATGGAGGAGG TGTCTGAGTA TCTTCAGCAA GTAGAAAATC AGTTGGAATC
51	CTGTTCCAAG CGATTAACCA AGATGGAAAC TTTTGCCCTTA GGTGTGAGGT
101	TGGAAGCTAA AGAAGAGATA GAGTCTATCA TACTTTCTGA TGTAGTGAAC
151	CGTTTGAGG TTTTATGTAG AGATATTGAA GATATGCTAT CTCGAGTCGA
201	GGAGATAGAG CGGATGTTAC GTATGGCGGA GCTTCCTCTA CTTCCCTATAA
251	AAGAACGCGCT TACCAAGGCT TTTGTACAAC ATAACAGCTG TAAAGAGAAG
301	TTAACCAAGG TAGAGCCTTA CTTTAAAGAG AGCCCTGCAT ATCTAACTAG
351	TGAAGAGCGA TTGCAGAGTT TGAATCAGAC TTTACAACGT GCGTACAAAG
401	AGTCCCCAAA GGTTTCAGGT TTAGAATCGG AAGTGGAGAGC CTGTCGAGAG
451	CAGCTTAAAG ATCAAGTAAG ACAGTTTGAA ACTCAAGGGAG TGAGCTTGAT
501	AAAAGAAGAG ATTCTCTTTG TGACTAGTAC CTTTAGAACT AAATTAGCT
551	ATCATTCCATT TCGATTACAT GTTCCCTTGCA TGAGGTTGTA TGAGGAGTAT
601	TATGATGACA TTGATCTAGA GAGAACTCGA GCTCGATGGA TGGCGATGTC
651	TGAGAGGTAT AGAGATGCTT TTCAGGCATT CCAGGAGATG TTGAAGGAAG
701	GCTTAGTGA AGAAGCTCAG GCTCTTAGAG AAACCGAGTA CTGGTTATAT
751	CGAGAGGAGA GAAAGAGTAA AAAGAAACAT TGA

35 The PSORT algorithm predicts a cytoplasmic location (0.523).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 49A.

The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 49B) and for FACS analysis (Figure 49C). A his-tagged protein was also expressed.

These experiments show that cp6296 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

40 Example 50

The following *C.pneumoniae* protein (PID 4376664) was expressed <SEQ ID 99; cp6664>:

1	MVLFHQAQASG RNRVKADAIV LPFWHKDAK NAASFEAEFE PSYLPALENF
51	QGKTGEIELL YSSPKAKEKR IVLLGLGKNE ELTSDVVVFQT YATLTRVLRK
101	AKCSTVNIIL PTISELRLSA EEFLVGLSSG ILSLNYYDPR YNKVDRNLET

151 PLSKVTVIGI VPKMADAIFR KEAAIFEGVY LTRDLVNRNA DEITPKKLAE
 201 VALNLGKEFP SIDTKVLGKD AIAKEKMGLL LAVSKGSCVD PHFIVVRYQG
 251 RPKSKDHTVL IGKGVTFDSC GLDLKPGKSM LTMKEDMAGG ATVLGILSAL
 301 AVLELPINVT GIIPATENAI DGASYKMGDV YVGMSGSLVE ICSTDAEGR
 351 ILADAITYAL KYCKPTRIID FATLTGAMVV SLGEEVAGFF SNNDVLAEDL
 401 LEASAETSEP LWRLPLVKKY DKTLMHSIAD MKNLGSNRAG AITAALFLQR
 451 FLEESSVAWA HLDIAGTAYH EKEEDRYPKY ASGFGVRSIL YYLENSLSK*

The cp6664 nucleotide sequence <SEQ ID 100> is:

1 GTGGTTTTAT TTCATGCTCA AGCCTCTGGG CGTAATCGTG TTAAGGCAGA
 10 51 TGCTATAGTC CTGCCCTTT GGCATTAA GGATGCAAAA AATGCAGCTT
 101 CTTTGAAAGC CGAGTTTGA CCCTCGTATC TCCCCGCTT AGAAAACCTT
 15 151 CAAGGAAAAA CCGGGGAGAT TGAACTCCTT TATAGTAGTC CTAAAGCTAA
 201 GGAAAAAACGC ATTGTCTCT TAGGCTTAGG GAAAAATGAA GAGCTCACCT
 251 CTGATGTGT TTTCCAAACC TATGCGACAC TAATCGTGT CTTACGTAAA
 301 GCAAAGTGTG CCACAGTCAA TATCATCTTA CCTACAATT CTGAATTGCG
 351 GCTTCTGCG GAAGAATTCT TAGTGGGGTT GTCCTCAGGA ATTTTGTAT
 401 TAAACTATGA CTACCCACGT TATAATAAGG TAGATCGTAA TCTTGAAACT
 451 CCTCTTCTA AAGTCACCGT TATCGGTATC GTTCCCAAA TGGCGGATGC
 501 TATCTTCTAGG AAAGAACAG CCATTTCGA AGGCGTATAT CTCACTCGAG
 551 ATCTTGTGAA CAGGAATGCT GATGAAATTAA CCCCTAAGAA ATTGGCAGAG
 601 GTTGCTCTGA ATCTGGGAAA AGAGTTCCCT AGTATTGATA CTAAGGTCTT
 651 GGGAAAAGAT GCCATCGCCA AAGAGAAAAT GGGACTCCCA TTGGCTGT
 701 CCAAGGGTTC TTGTGTTGGAT CCACACTTTA TCCTTGTCCG TTATCAAGGA
 751 CGTCCTAAGT CTAAAGATCA CACCGTCTTG ATAGGAAAG GGGTCACTTT
 801 TGACTCTGGA GTTTAGACC TCAAGCCTGG AAAATCCATG CTTACTATGA
 851 AAGAAGACAT GGCAGGTGGG GCTACAGTC TCAGGGATTCT CTCGGCGTTA
 901 GCAGTTTAG AGCTTCCTAT AAATGTCACG GGGATCATTC CTGCTACAGA
 951 GAATGCTATC GATGGCCCT CCTATAAAAT GGGAGATGTC TATGTAGGAA
 1001 TGTCGGGGCT TTCTGTTGAG ATTGTAGTA CCGATGCTGA GGGACGTCTT
 1051 ATCCTCGCTG ATGCGATTAC ATATGCTTA AAATATTGTA AACCGACACG
 1101 TATTATAGAT TTTGCAACTC TAACAGGAGC TATGGTAGTC TCTCTAGGAG
 1151 AAGAGGTTGC AGGTTCTTT TCCAATAACG ATGTTTTAGC TGAAGATCTT
 1201 TTAGAGGCGT CAGCCGAAAC CTCCGAGCCG TTATGGAGAC TTCCCTCTAGT
 1251 TAAGAAGTAT GATAAAACAT TGCAATTCTGA TATTGCTGAT ATGAAAATC
 1301 TAGGCAGTAA CCGTGCAGGG GCTATTACAG CAGCATTATT CTTGCAGAGA
 1351 TTTTTGGAAG AATCTTCGGT AGCTTGGGCA CATCTTGATA TTGCAGGTAC
 1401 TGCAATATCAT GAAAAAGAAG AAGACCCTTA TCCAAAATAT GCTTCAGGTT
 1451 TTGGTGTTCG TTCTATTCTT TATTACTTAG AAAATAGTCT TTCTAAAGTAG

The PSORT algorithm predicts an inner membrane location (0.268).

40 The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 50A), as a his-tagged protein, and as a GST/His fusion. The proteins were used to immunise mice, whose sera were used in Western blot Western blot (50B) and FACS (50C) analyses.

The cp6664 protein was also identified in the 2D-PAGE experiment (Cpn0385) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

45 These experiments show that cp6664 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 51

The following *C.pneumoniae* protein (PID 4376696) was expressed <SEQ ID 101; cp6696>:

50 1 MTLIFVIIIV WCNAFLIKLC VIMGLQSRLQ HCIEVSQNSN FDSQVKQFIY
 51 ACQDKTLRQS VLKIFRYHPL LKIHDIAARAV YLLMALEEDE DLGLSFLNVQ
 101 QYPSGAVELF SCGGFPWKGL PYPAEHAEFG LLLLQIAEFY EESQAYVSKM
 151 SHFQQALFDH QGSVFPSSLWS QENSRLLKEK TTLSQSFLFQ LGMQIHPEYS
 201 LEDPALGFWM QRTRSSSAFV AASGCQSSLG AYSSGDVGVI AYGPCSGDIS
 251 DCYYFGCCGI AKEFVCQKSH QTTEISFLTS TGKPHPRNTG FSYLRDSYVH
 301 LPIRCKITIS DKQYRVHAAL AEATSAMTF5 IFCKGKNCQV VDGPRLRSCS

-91-

351 LDSYKGPGND IMILGENDAI NIVSASPyme IFALQGKEKF WNADFLINIP
 401 YKEEGVMLIF EKKVTSEKGR FFTKMN*

A predicted signal peptide is highlighted.

The cp6696 nucleotide sequence <SEQ ID 102> is:

```

5      1 TTGACTCTAA TTTTGTTAT TATTATCGTT TGGTGCATG CTTTCTGAT
      51 CAAATTGTGC GTGATAATGG GGCTGCAATC CAGGTTACAA CATTGTATAG
     101 AAGTGTCCA GAATTCGAAC TTTGATTAC AAGTAAAACA GTTTATCTAT
     151 GCGTGCAGA ATAAGACATT AAGGCAGTCT GTACTCAAGA TTTTCCGCTA
    201 CCATCCTTTA CTAAAAATTC ATGATATTGC TCAGGCCGTC TATCTTTGA
    251 TGGCCTTAGA AGAAGGGAG GATTAGGCT TAAGCTTTT AAATGTACAG
    301 CAGTACCCCT CAGGTGCTGT AGAACATGTT TCTTGTGGGG GATTTCCCTG
    351 GAAAGGATTA CCTTATCCTG CAGAACATGC GGAATTGGC CTACTCCGT
    401 TACAGATCGC AGAGTTTAT AGAGAGAGTC AGGCATACGT CTCTAAAATG
    451 AGTCATTTTC AACAGGCCT CTTTGATCAC CAAGGGAGCG TCTTCCCTC
   15 501 TCTCTGGAGC CAGGAGAATC CTCGACTCC AAAAGAAAAG ACAACTCTTA
   551 GCCAATCGTT TCTCTTCAA TTAGGAATGC AAATTCAACC AGAATACAGT
   601 CTTGAGGATC CTGCACTAGG GTTCTGGATG CAAAGAACGC GTTCTTCATC
   651 CGCTTTGTGTA GCGCCTTCAG GATGCAAAG TAGCTTGGGA GCGTATTCC
   701 CAGGGGATGT CGGTGTTATC GCTTATGGAC CTTGCTCTGG AGACATTAGT
   751 GATTGTTATT ATTTTGGATG TTGTTGGAATC GCTAAAGAGT TCGTGTGCCA
   801 AAAATCTCAC CAAACTACAG AGATTTCTTT TCTCACCTCT ACAGGAAAGC
   851 CTCATCCCAG AAATACGGGA TTTTCTTACCC TTGAGATTC CTATGTACAT
   901 CTGCCGATCC GCTGTAAGAT CACTATTCC GACAAGCAAT ATCGCGTGCA
   951 CGCTGCGTTG GCTGAGGGCA CCTCTGCCAT GACGTTTCT ATTTCTGTA
  25 1001 AGGGGAAGAA TTGTCAGGTT GTTGACGGCC CTCGCTTGGC CTCCTGTTCC
  1051 CTAGATTCTT ATAAAGGTCC CGGAAACGAC ATTATGATTC TTGGGGAAAAA
  1101 TGACGCAATC AACATTGTT CTGCAAGTCC CTATATGGAA ATTTTGCTT
  1151 TGCAAGGCAA AGAAAAATTG TGGAAATGCAG ACTTTTGAT TAATATTCC
  1201 TACAAAGAAG AGGGCGTCAT GTTAATTGTTT GAAAAAAAAG TGACCTCTGA
  30 1251 GAAAGGAAGA TTCTTTACGA AGATGAATTA A

```

The PSORT algorithm predicts an inner membrane location (0.463).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 51A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 51B) and for FACS analysis (Figure 51C). A his-tagged protein was also expressed.

35 This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6696 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 52

40 The following *C.pneumoniae* protein (PID 4376790) was expressed <SEQ ID 103; cp6790>:

```

1 MSEHKKSSKI IGIDLGTNS CVSVMEGGQA KVITSSEGTR TPPSIVAFKG
51 NEKLVGIPAK RQAVTNPEKT LGSTKRFIGR KYSEVASEIQ TVPYTVTSGS
101 KGDAVFEVDG KQYTPEEIGA QILMKMKETA EAYLGETVTE AVITVPAYFN
151 DSQRASTKDA GRIAGLDVKR IIPEPTAAAL AYGIDKVGDK KIAVFDLGGG
201 TFDISILEIG DGVFEVLSTN GDTLLGGDDF DEVIIKWMIE EFKKQEGIDL
251 SKDNMALQRL KDAAEKAKIE LSGVSSTEIN QPFITMDAQG PKHLALTLTR
301 AQFEKLAASL IERTKSPCIK ALSDAKLSAK DIDDVLLVGG MSRMPAVQET
351 VKELFGKEPN KGVPDDEVVA IGAAIQGGVL GGEVKDVLLL DVIPLSLGIE
401 TLGGVMTTLV ERNTTIPTQK KQIFSTAADN QPAVTIVVLO GERPMKDKN
451 EIGRFDLTDI PPAPRGHPQI EVSFIDANG IFHVSAKDVA SGKEQKIRIE
501 ASSGLQEDEI QRMVRDAEIN KEEDKKRREA SDAKNEADSM IFRAEKAIKD
551 YKEQIPELTV KEIEERIENV RNALKDDAPI EKIKEVTELD SKHMQKIGES
601 MQSQSASAAA SSAANAKGGP NINTEDLKKH SFSTKPPSNN GSSEDHIEEA

```

651 DVEIIDNDDK*

The cp6790 nucleotide sequence <SEQ ID 104> is:

```

5      1 ATGAGTGAAC ACAAAAATC AAGCAAATT ATAGGTATAG ACTTAGGCAC
      51 AACAAACTCC TCGGTATCTG TTATGGAAGG AGGACAAGCT AAAGTAATTA
     101 CATCATCCGA AGGAACAAGA ACCACGCCAT CGATCGTTGC CTTCAAAGGT
     151 AATGAGAAAT TAGTGGGAT TCCAGCAAA CGTCAAGCAG TGACAAATCC
    201 AGAAAAAAACT CTCGGCTCTA CAAAACGCTT TATTGGCCGT AAGTACTCTG
    251 AAGTAGCTTC GGAAATCCAA ACCGTTCCCTT ATACAGTCAC CTCCGGATCT
    301 AAAGGTGATG CCGTTTCGA AGTTGATGGC AAACAATACA CTCCAGAAGA
    351 AATTGGCGCA CAAATCTTAA TGAAAATGAA AGAGACAGCA GAAGCTTATC
    401 TAGGCGAAAC TGTCACAGAA GCAGTGATCA CCGTCCCCGC ATACTTCAAT
    451 GATTCTCAAC GAGCATCCAC AAAAGATGCT GGACGCATTG CAGGTCTAGA
    501 TGTAAAACGT ATCATTCCAG AACCTACCAGC AGCAGCTCTT GCCTACGGAA
    551 TCGATAAAAGT CGGTGATAAAA AAAATCGCTG TCTTCGACCT TGGTGGAGGA
    601 ACTTTTGATA TCTCCATCCT AGAAATCGGT GATGGCGTCT TCGAAGTTCT
    651 ATCTACAAAT GGAGATACTC TCCTCGGTGG AGACGACTTT GATGAAGTCA
    701 TTATCAAATG GATGATCGAA GAATTCAAAA ACAAGAAAGG CATTGATCTT
    751 AGCAAAGATA ATATGGCCTT ACAAAAGACTT AAAGATGCTG CTGAGAAAGC
    801 AAAAATAGAA CTTTCAGGAG TCTCTTCCAC AGAAATCAAT CAGCCATTCA
    851 TCACAATGGG TGCAACAAGG CCTAAACACC TTGCATTGAC ACTCACACGT
    901 GCGCAATTCC AGAAACTCGC AGCCTCTCTA ATCGAAAGAA CAAAATCTCC
    951 ATGCATCAAAGA GCACTCAGTG ACGCAAAGT TTCCGCTAAG GATATCGATG
   1001 ATGTTCTCTT AGTTGGAGGT ATGTCAGAA TGCCCGCAGT GCAAGAAACT
   1051 GTAAAAGAAC TCTTCGGCAA AGAGCCTAAT AAAGGAGTC ACCCCGACGA
   1101 AGTTGTTGCT ATTGGAGCCG CAATTCAAGG TGGTGTCTT GCGGAGAAAG
   1151 TTAAGGATGT TCTACTCTCA GACGTATCTC CCCTATCTCT GGTATCGAA
   1201 ACTCTAGGAG GCGTCATGAC GACTCTGGTA GAGAGAAATA CTACAATCCC
   1251 TACACAGAAA AAACAAATCT TCTCCACAGC TGCTGATAAC CAGCCTGCAG
   1301 TTACCATCGT AGTTCTCCAA GGAGAGCGTC CCATGGCCAA AGATAACAAG
   1351 GAAATCGGAA GATTCGATCT TACAGATATC CCTCCGGCTC CTCGAGGCCA
   1401 TCCTCAAATC GAAGTCTCCT TCGATATCGA TGCAAAACGGA ATTTCATG
   1451 TCTCAGCTAA AGATGTTGCC AGCGGTAAAG AACAGAAAAT TCGTATCGAA
   1501 GCAAGCTCAG GACTTCAGA AGATGAAATC CAAAGAATGG TTCGAGATGC
   1551 CGAAATTAAAT AAGGAAGAAG ATAAAAAAAGC TCGTGAAGCT TCAGATGCTA
   1601 AAAATGAAGC CGATAGCATG ATCTTCAGAG CCGAAAAAGC TAITAAAGAT
   1651 TATAAGGAGC AAATTCTGA AACTTTAGTT AAAGAAATCG AAGAGCGAAT
   1701 CGAAAACGTG CGCAACGCAC TCAAAGATGA CGCTCCTATT GAAAAAAATTA
   1751 AAGAGGTTAC TGAAGACCTA AGCAAGCATA TGCAAAAAAT TGGAGAGTCT
   1801 ATGCAATCGC AGTCTGCATC AGCAGCAGCA TCATCGGCAG CCAATGCTAA
   1851 AGGTGGACCT AACATCAATA CAGAAGATTT GAAAAAAACAT AGTTTCAGTA
   1901 CGAAGCCTCC TTCAAATAAC GGTTCTTCAG AAGACCATAT CGAAGAAGCT
   1951 GATGTAGAAA TTATTGATAA CGACGATAAG TAA

```

The PSORT algorithm predicts an inner membrane location (0.151).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 52A) and a his-tagged product. The proteins were used to immunise mice, whose sera were used in Western blot (Figure 52B) and FACS (Figure 52C) analyses.

The cp6790 protein was also identified in the 2D-PAGE experiment (Cpn0503).

These experiments show that cp6790 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

50 Example 53

The following *C.pneumoniae* protein (PID 4376878) was expressed <SEQ ID 105; cp6878>:

```

55      1 MNVPDSKNLH PPAYELLEIK ARITQSYKEA SAILTAIPDG ILLSETGHF
      51 LICNSQAREI LGIDENLEIL NRSFTDVLPD TCLGFSIQEA LESLKVPKTL
     101 RLSLCESKE KEVELFIRKN EISGYLFIQI RDRSDYKOLE NAIERYKNIA
     151 ELGKMTATLA HEIRNPLSGI VGFASILKKE ISSPRHQRLM SSIISGTRSL
     201 NNLVSSMLEY TKSQPLNLKI INLQDFFSSL IPILLSVSFPN CKFVREGAQP

```

251 LFRSIDPDRM NSVWNLVKN AVETGNSPIT LTLHTSGDIS VTNPGTIPSE
 301 IMDKLFTPFF TTKREGNGLG LAEAQKIIRL HGGDIQLKTS DSAVSFFIII
 351 PELLAALPK E RAAS*

The cp6878 nucleotide sequence <SEQ ID 106> is:

5 1 ATGAACGTCC CTGATTCCAA GAACCTCCAT CCTCCCTGCAT ACGAAACTCCT
 51 AGAGATCAAG GCTCCCATCA CACAATCTTA TAAAGAAGCG AGTGCTATAAC
 101 TGACAGCGAT TCCCTGATGGT ATCCTATTAC TTTCTGAAAC AGGACACTTT
 151 CTTATCTGCA ATTACAAAGC ACGTGAAATT CTAGGAATTG ATGAAAATCT
 201 AGAAAATTCTT AATAGATCCT TTACCGATGT TCTCCCGAT ACGTGTCTTG
 251 GATTTTCAT TCAAGAGGCT CTTGAATCTC TAAAAGTCCC TAAAAGTCTT
 301 AGACTCTCTC TCTGTAAAAGA ATCTAAAGAA AAAGAAGTGG AACTCTTCAT
 351 CCGTAAAAAC GAGATCAGTG GATACCTGTT TATCCAATC CGCGATCGGT
 401 CCGACTATAA ACAACTAGAA AACGCTATAG AAAGATATAA AAATATCGCA
 451 GAACTTGGGA AAATGACGGC TACCCTAGCT CACGAAATCC GCAATCCGCT
 501 AAGTGGAAATC GTTGGATTTG CCTCTATCCT AAAGAAAGAG ATTTCCCTCTC
 551 CTCGCCACCA ACGAATGCTC TCCTCAATCA TCTCCGGCAC AAGGTCTCTA
 601 AATAACCTTG TCTCTTCTAT GTTGAATAT ACAAAATCAC AACCGTTGAA
 651 CCTAAAGATT ATAAATTTCAG AAGACTTCTT CTCTCTCTT ATCCCTCTGC
 701 TCTCCGTCTC TTTCCCGAAT TGCAAGTTG TAAGAGAGGG CGCACAAACCT
 751 CTATTCAAGAT CTATAGATCC TGATCGGATG AACAGTGTG TTTGGAACCT
 801 AGTAAAAAT GCTGTAGAAA CAGGGAACTC TCCGATCACT CTGACCCCTGC
 851 ATACATCGGG AGACATCTCG GTAACGAACC CGGGAAACGAT TCCTTCCGAG
 901 ATCATGGACA AGCTCTTCAC TCCATTCTC ACAACAAAGA GAGAGGGAAA
 951 TGGTTTGGGA CTTGCTGAAG CTCAAAAT TATAAGACTC CATGGAGGAG
 20 1001 ATATCCAATT AAAAACAAAGC GACTCCGCG TTAGCTTCTT CATAATCATC
 1051 CCCGAACCTTC TAGCGGCCCT ACCCAAAGAA AGAGCCGCTA G

The PSORT algorithm predicts an inner membrane location (0.204).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 53A) and as a GST-fusion product. The recombinant GST-fusion protein was used to immunise mice, whose sera were 30 used in a Western blot (Figure 53B) and for FACS analysis.

These experiments show that cp6878 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 54

The following *C.pneumoniae* protein (PID 4377224) was expressed <SEQ ID 107; cp7224>:

35 1 MMKKIRKVAL AVGGSGGHIV PALSVKEAFS REGIDVLLLG KGLKNHPSLQ
 51 QGISYREIPS GLPTVLNPPIK IMSRTLSLCS GYLKARKELK IFDPDLVIGF
 101 GSYHSLPVLL AGLSHKIPLF LHEQNLVPGK VNQLFSRYAR GIGVNFSPTV
 151 KHFRCPAEEV FLPKRFSFLG SPMMKRCTNH TPTICVVGGS QGAQILNTCV
 201 PQALVVKLVNK YPNMYVHHIV GPKSDVMKVQ HVYNRGEVLC CVKPFEEQLL
 40 251 DVLLAADLVI SRAGATILEE ILWAKVPGIL IPYPGAYGHQ EVNAKFFVDV
 301 LEGGTMILEK ELTEKLLVEK VTFALDSHNR EKQRNSLAAY SQQRSTKTFH
 351 AFICECL*

The cp7224 nucleotide sequence <SEQ ID 108> is:

45 1 ATGATGAAGA AAATTCGAAA AGTAGCCTTG GCTGTAGGAG GTTCAGGAGG
 51 CCACATTGTC CCAGCTCTCT CGGTAAAGGA AGCTTTTTCT CGTGAAGGAA
 101 TAGACGTATT ACTACTAGGG AAAGGTCTCA AGAACCATCC TTCTTTGCAA
 151 CAGGGAATCA GCTATCGGA AATCCCTCA GGACTTCCTA CAGTCCTTAA
 201 TCCCATAAAAG ATCATGAGCA GGACCCCTTC TCTATGTTCA GGATAACCTGA
 251 AAGCAAGAAA GGAACCTAAA ATTTCGACCC CTGACCTGGT CATAGGATTT
 50 301 GGGAGCTACC ACTCTCTTCC CGTGTGCTC GCAGGACTGT CCCATAAAAT
 351 TCCCTTATTCT CTACACGAAC AAAATCTAGT TCCTGGAAA GTAATCAAT
 401 TGTTTTCCCG CTATGTCGA GGTATTGGAG TGAATTCTC CCCCCCTTA
 451 AAACACTTCC GCTGCCCG AGAAGAGGTC TTCCCTCCTA AACGAAGCTT
 501 CTCTCTTAGGA AGCCCTATGA TGAAGCGATG TACAAATCAT ACCCCTACAA
 55 551 TCTGTGTTGT TGGAGGTCT CAGGGAGCAC AGATATTAAG TACTTGTGTT
 601 CCCCCAGCTC TTGTCAAGCT AGTCAATAAG TACCCAAATA TGTACGTCCA

651 TCATATTGTA GGACCTAAAA GTGATGTTAT GAAGGTGCAA CATGTTACA
 701 ATCGTGGAGA GGTCCCTGTC TGTGTGAAGC CGTTCGAAGA GCAACTCCTA
 751 GATGTCTTGC TTGCCGAGA TTTGGTCATC AGTAGGGCAG GAGCCACAAT
 801 TTAGAAGAA ATTCTTTGGG CAAAAGTTCC CGGAATTATA ATTCCCTATC
 851 CAGGAGCTTA TGGACATCAG GAAGTTAATG CTAAATTCTT TGTAGACGTC
 901 TTAGAAGGGG GAACTATGAT CCTAGAAAAA GAATTAACAG AGAAGCTATT
 951 AGTAGAAAAA GTAACGTTG CTTAGACTC CCATAACAGA GAAAACAAC
 1001 GCAATTCCCT AGCGGCGTAT AGTCAGCAAA GGTCAACAAA AACATTCCAT
 1051 GCATTCATTT GTGAATGCTT ATAG

- 10 The PSORT algorithm predicts an inner membrane location (0.164).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 54A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 54B) and for FACS analysis (Figure 54C). A his-tagged protein was also expressed.

- 15 This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7224 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 55

The following *C.pneumoniae* protein (PID 4377140) was expressed <SEQ ID 109; cp7140>:

20 1 **MVRSSISFCL FFLMTLICCT SCNSRSLIVH GLPGREANEI VVLLVSKGVVA**
 51 51 **AQKLPQAAAAA** TAGAATEQMW DIAVPSAQIT EALAILNQAG LPRMKGTSSL
 101 101 DLFAKQGLVP SELQEKIRYQ EGLSEQMAST IRKMDGVVDA SVQISFTTEN
 151 151 EDNLPLTASV YIKHRGVLND PNSIMVSKIK RLIASAVPGL VPENVSVVSD
 201 201 RAAYSGITIN GPWGLTEEID YVSVWGIILA KSSLTKFRLI FYVLILILFV
 25 251 ISCGLLWVIW KTHTLIMTMG GTKGFFNPTP YTNALEAKK AEGAADKEK
 301 301 KEDADSQGES KNAETSDKDS SDKDAPEGSN EIEGA*

A predicted signal peptide is highlighted.

The cp7140 nucleotide sequence <SEQ ID 110> is:

30 1 ATGGTTCGTC GATCTATTTTC TTTTTGCTTG TTCTTTCTAA TGACATTGCT
 51 51 GTGCTGTACA AGCTGTAAACA GCAGGTCTCT AATTGTGCAC GGTCTTCCTG
 101 101 GCAGAGAACG GAATGAGATT GTGGTCTTT TGGTAAGCAA AGGGGTGGCT
 151 151 GCACAAAAT TGCCTCAAGC TGCGAGGGCT ACAGCCGGAG CAGCTACTGA
 201 201 GCAAATGTGG GATATCGCGG TTCCGTCAAGC ACAAAATCACA GAGGCCCTTG
 251 251 CCATTCTAAA TCAAGCGGGT CTTCCACGTA TGAAAGGGAC AAGCCTGTTA
 301 301 GATCTTTTTG CAAAACAAGG TCTTGTCTT TCCGAGCTTC AGGAAAAAAAT
 35 351 CCGTTATCAA GAAGGCTTAT CAGAACAGAT GGCCTCTACG ATTAGAAAAA
 401 401 TGGATGGCGT TGTGATGCC TCAGTACAGA TTTCCTTCAC TACAGAAAAT
 451 451 GAAGATAATC TTCCCTTAAC AGCCTCTGTG TATATTAAGC ATCGAGGGT
 501 501 TTTGGACAAT CCGAACAGCA TTATGGTTTC CAAAATTAAG CGCCTTATTG
 551 551 CAAGTGTGT TCCAGGACTT GTGCCAGAGA ACGTCTCTGT AGTGAGCGAT
 601 601 CGCGCAGCTT ATAGTGATAT TACAATTAAT GGTCTTGGG GATTAACAGA
 651 651 AGAAATCGAT TATGTTCTG TTTGGGGTAT TATTCTGCG AAGTCTTCGC
 701 701 TCACCAAATT CCGTCTCATT TTTATGTCT TGATTCTCAT TTTATTTGTT
 751 751 ATTCTTGTG GTCTCCTTTG GGTCAATTGG AAAACTCATA CTCTCATTAT
 801 801 GACTATGGGA CGTACAAAAG GGTCTTCAA CCCTACACCA TATACAAAGA
 851 851 ATGCCCTTGGGA AGCCAAGAAA GCCGAGGGAG CAGCTGCTGA CAAAGAGAAA
 901 901 AAAGAAGATG CAGATTCAAA GGGGGAAAGC AAAATGCCG AAACCAGTGA
 951 951 TAAAGACTCT AGTGATAAAAG ATGCTCCAGA AGGAAGCAAT GAAATTGAGG
 1001 1001 GTGCTTAG

- 50 The PSORT algorithm predicts an inner membrane location (0.650).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 55A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 55B) and for FACS analysis (Figure 55C). A his-tagged protein was also expressed.

These experiments show that cp7140 is a surface-exposed and immunoaccessible protein, and that it
5 is a useful immunogen. These properties are not evident from the sequence alone.

Example 56

The following *C.pneumoniae* protein (PID 4377306) was expressed <SEQ ID 111; cp7306>:

10	1 MITKQLRSWL AVLVGSSLLA LPLSGQAVGK KESRVSELPQ DVLLKEISGG
	51 FSKVATKATEP AVVYIESFPK SQAVENTPSPG RRGPYENPFD YFNDEFFNRF
	101 FGLPSQREKP QSKEAVRGTG FLVSPDGYIV TNNHHVEDTG KIHVTLHDGQ
	151 KYPATVIGLDPKTDLAVIKI KSQNLPYLSF GNSDHLKVGD WAIAGNPG
	201 LQATVTVGVI SAKGRNQLHI ADFEDFIQTD AAINPGNSGG PLLNIDGQVI
	251 GVNTAIVSGS GGYIGIGFAI PSLMANRIID QLIRDGQVTR GFLGVTLQPI
15	301 DAEALAAKYKL EKVY GALVTD VVKGSPADKA GLKQEDVIIA YNGKEVDSL
	351 MPRNAVSLMN PDTRIVLKVV REGKVIEIPV TVSQAPKEDG MSALQRVGIR
	401 VQNLTPETAK KLGIAPETKG ILIISVEPGS VAASSGIAPG QLILAVNRQK
	451 VSSIEDLNRT LKDSNNENIL LMVSQGDVIR FIALKPEE*

A predicted signal peptide is highlighted.

The cp7306 nucleotide sequence <SEQ ID 112> is:

20	1 ATGATAACTA AGCAATTGCG TTCGGGCTA GCTGTACTTG TTGGTTCAAG
	51 TCTGCTAGCT CTTCTTAT CAGGGCAAGC TGTCGGGAAA AAAGAACATCTC
	101 GAGTTCCGA GCTGCCTCAA GACGTTCTTC TTAAAGAGAT CTCGGGAGGG
	151 TTTCTAAGG TCGCTACCAA GGCAGACTCCC GCTGTTGTGT ACATAGAAAG
	201 TTTCCCAAAG AGCCAGGCTG TAACACATCC TTCTCCTGGA CGCCGTGGC
25	251 CTTATGAAAAA TCCTTTTGAT TATTTAATG ATGAGTTTT CAATCGTTT
	301 TTTGGTCTAC CTTCACAGAG GGAAAAACCT CAAAGTAAG AGGCGGTTCG
	351 AGGAACAGGT TTCCCTAGTAT CTCCAGATGG CTATATTGTG ACTAATAACC
	401 ATGTTGTCGA AGATACAGGT AAGATTCAAG TAACTCTICA TGATGGCAA
30	451 AAGTACCCAG CAACTGTAAT CGGACTCGAT CCTAAAACAG ACCTTGCAGT
	501 CATTAAAATT AAATCCCCAA ACCTCCCGTA TCTTTCTTTT GGAAACTCCG
	551 ACCACTTAAA AGTCGGAGAT TGGCAATTG CAATTGGAAA TCCCTCGGT
	601 CTTCAAGCTA CGGTACCGT AGGTGTCATC AGTGTCAAAG GAAGAAATCA
	651 ACTCCACATT GCAGATTTG AAGATTTAT TCAGACAGAT GCTGGATTA
	701 ATCCAGGCAA CTCTGGAGGC CCTCTTCTAA ATATGATGG ACAGGTCATC
35	751 GGTGTTAATA CTGCCATTGT CAGTGGTAGT GGTGGCTATA TTGGAATCGG
	801 GTTTCGATT CCTAGCCTTA TGGCAAATAG AATCATAGAT CAGCTGATTC
	851 GTGATGGTCA AGTTACCGA GGATTCTTAG GAGTGTACTT ACAACCTATA
	901 GATGCGGAAC TCGCTGTTG CTACAAACTC GAAAAGTTT ATGGCGCTTT
	951 AGTCACAGAT GTTGTAAAG GATCTCCAGC AGATAAAAGCA GGGCTAAAAC
40	1001 AAGAAGATGT GATCATTGCT TATAATGGGA AAGAAGTCGA TTCACTGAGT
	1051 ATGTTCCGTA ATGCTGTTTC TTTAATGAAT CCAGATACAC GTATTGTTCT
	1101 AAAGGTAGTT CGTGAAGGAA AGTTATCGA AATACCCGTG ACAGTTTCTC
	1151 AAGCTCCAAA AGAAGATGGA ATGTCGGCTT TACAGCGTGT GGGATCCGT
	1201 GTGCAAAACC TAACTCCTGA AACTGCTAAG AAGCTGGAA TTGCTCCAGA
45	1251 GACTAAAGGC ATTTTGATTA TAAGTGTGTA ACCAGGGTCT GTAGCAGCTT
	1301 CTTCAAGGAAT TGCTCCTGGT CAGCTGATCC TTGCTGTGAA TAGACAAAAA
	1351 GTATCTTCGA TTGAAGATCT GAATAGAACG TTAAAGATT CTAACAATGA
	1401 GAATATTCTT CTTATGGTTT CTCAAGGAGA TGTTATTCCG TTCATTGCC
	1451 TGAAACCTGA AGAATAA

50 The PSORT algorithm predicts a periplasmic location (0.923).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 56A) and as a GST-fusion product (Figure 56B). The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 56C) and for FACS (Figure 56D) analyses.

The cp7306 protein was also identified in the 2D-PAGE experiment (Cpn0979) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7306 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

5 Example 57

The following *C.pneumoniae* protein (PID 4377132) was expressed <SEQ ID 113; cp7132>:

```

1 MCNSTIAMKKQ KRGFVLMELL MSFTLIALLL GTLGFWYRKI YTVQKQKERI
51 YNFYIEESRA YKQLRTLFSM SLSSSYEEPG SLFSLIFDRG VYRDPKLAGA
101 VRASLHHDTK DQRLELRICN IKDQSYFETQ RLLSHVTHVV LSFQRNPDPDE
151 KLPETIALTI TREPKAYPPR TLTYQFAVGK*

```

A predicted signal peptide is highlighted.

The cp7132 nucleotide sequence <SEQ ID 114> is:

```

1 ATGTGTAACT CTATAGCTAT GAAAAAGCAA AAGCGTGGCT TTGTGCTTAT
51 GGAATTACTC ATGTCGTTCA CTCTAATTGC TTTGTTATTA GGGACTTTAG
101 GATTTTGGTA TCGGAAATT TATACTGTAC AAAAGCAAAA AGAACGTATT
151 TATAACTTTT ATATCGAAGA AAGCCGAGCC TACAAGCAGC TCAGAACCCCT
201 GTTTAGCATG TCCTTGTCCT CATCTTACGA GGAGCCTGGA TCATTATTTT
251 CTTTAATCTT TGATCGGGGT GTTATCGAG ATCCTAACGCT GGCAGGTGCG
301 GTACGAGCTT CTCTCCATCA TGACACCAAG GATCAGAGAT TGGAACCTCG
351 TATTTGTAAT ATTAAGGATC AGTCTTACTT TGAAACACAG CGACTGCTCT
401 CCCACGTGAC CCATGTTGTA CTTCCCTTCC AGAGAAATCC TGATCCTGAA
451 AAACCTCCCTG AAACAATTGC TTTAACTATA ACACGGGAAC CTAAGCATA
501 TCCTCCAAGG ACGTTAACAT ACCAATTGTC GGTTGGGAAA TAA

```

The PSORT algorithm predicts a periplasmic location (0.915).

25 The protein was expressed in *E.coli* and purified as a his-tag product (Figure 57A) or as a GST-fusion. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 57B) and FACS (Figure 57C) analyses.

These experiments show that cp7132 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

30 Example 58

The following *C.pneumoniae* protein (PID 4376733) was expressed <SEQ ID 115; cp6733>:

```

1 MKTSIPWVLV SSVLAFSCHL QSLANEELLS PDDSFNGNID SGTFTPCKTSA
51 TTYSLTGDFV FYEPGKGTPQ SDSCFKQTTD NLTFLGNGHS LTFGFIDAGT
101 HAGAAAASTTA NKNLTFSGFS LLSDFDSSPST TVITFGQGTLs SAGGVNLENI
151 RKLVVAGNFS TADGGAIKGA SFLLTGTSGD ALFSNNSSST KGGAIATTAG
201 ARIANNTGYV RFLSNIASTS GGAIDDEGTS ILSMNKFLYF EGNAAKTTGG
251 AICNTKASGS PELIISNNKT LIFASNVAET SGGAIAHAKL ALSSGGFTEF
301 LRNNVSSATP KGGAIISIDAS GELSLSAETG NITFVRNTLT TTGSTDTPKR
351 NAINIGNSNGK FTELRAAKNH TIFFYDPITS EGTSSDVLIK NNGSAGALNP
401 YQGTILFSGE TLTADELKVA DNLKSSFTQP VSLSGGKLLL QKGVTLESTS
451 FSQEAGSLLG MDSGTTLSTT AGSITITNLG INVDSLGLKQ PVSLTAKGAS
501 NKVIVSGKLN LIDIEGNIYE SHMFSHDQLF SLLKITVDAD VDTNVDISSL
551 IPVPAEDPNS EYGFQGQWNV NWTTDTATNT KEATATWTKT GFVPSPERKS
601 ALVCNTLWGV FTDIRSLQQL VEIGATGMEH KQGFVWSSMT NFLHKTGDEN
651 RKGRFRHTSGG YVIGGSAHTP KDDLFTFAFC HLFARDKDCF IAHNNSRTYG
701 GTLFFKHSHT LQPQNYRLG RAKFSESAIE KFPREIPLAL DVQVSFSHSD
751 NRMETHYTSI PESEGWSNE CIAGGIGLDL PFVLSNPHPPL FKTFIPQMVK
801 EMVYVSQNSF FESSSDGRGF SIGRLLNLSI PVGAKFVQGD IGDSYTYDLS

```

-97-

851 GFFVSDVYRN NPQSTATLVM SPDSWKIRGG NLSRQAFLLR GSNNYVYNSN
 901 CELFGHYAME LRGSSRNYNV DVGTKLRF*

A predicted signal peptide is highlighted.

The cp6733 nucleotide sequence <SEQ ID 116> is:

5	1	ATGAAGACTT CGATTCCCTG GGTTTAGTT TCCTCCGTGT TAGCTTCCTC
	51	ATGTCACCTA CAGTCACTAG CTAAACGAGGA ACTTTTATCA CCTGATGATA
	101	GCTTTAATGG AAATATCGAT TCAGGAACGT TTACTCCAAA AACTTCAGCC
	151	ACAACATATT CTCTAACAGG AGATGTCTTC TTTTACGAGC CTGGAAAAGG
10	201	CACTCCCTTA TCTGACAGTT GTTTAACGCA AACACGGAC AATCTTACCT
	251	TCTTGGGAA CGGTCTAGC TTAACGTTTG GCTTTATAGA TGCTGGCACT
	301	CATGCAGGTG CTGCTGCATC TACAACAGCA ATAAGAACAT TTACCTTCCTC
	351	AGGGTTTCC TTACTGAGTT TTGATTCCCTC TCCTAGCACA ACGGTTACTA
	401	CAGGTCAAGG AACGCTTTCC TCAGCAGGAG GCGTAAATT AGAAAATATT
	451	CGTAAACTTG TAGTTGCTGG GAATTTCCTT ACTGCAGATG GTGGAGCTAT
15	501	CAAAGGAGCG TCTTTCCCTT TAACTGGCAC TTCTGGAGAT GCTCTTTTA
	551	GTAACAACTC TTCATCAACA AAGGGAGGAG CAATTGCTAC TACAGCAGGC
	601	GCTCGCATAG CAAATAACAC AGGTTATGTT AGATTCCAT CTAACATAGC
	651	GCTCTACGTCA GGAGGCCCTA TCGATGATGA AGGCACGTCG ATACTATCGA
20	701	ACAACAAATT TCTATATTTT GAAGGGAATG CAGCGAAAAC TACTGGCGGT
	751	GCGATCTGCA ACACCAAGGC GAGTGGATCT CCTGAACCTGA TAATCTCTAA
	801	CAATAAGACT CTGATCTTGT CTTCAAAACGT AGCAGAAACA AGCGGTGGCG
	851	CCATCCATGC TAAAAAGCTA GCCCTTCCT CTGGAGGCTT TACAGAGTTT
	901	CTACGAAATA ATGTCTCATC AGCAACTCCT AAGGGGGGTG CTATCAGCAT
	951	CGATGCCCTCA GGAGAGCTCA GTCTTCTGC AGAGACAGGA AACATTACCT
25	1001	TTGTAAGAAA TACCCTTACA ACAACCGGAA GTACCGATAC TCCTAAACGT
	1051	AATGCGATCA ACATAGGAAG TAACGGGAAA TTACCGGAAT TACGGGCTGC
	1101	TAAAAATCAT ACAATTTCCT TCTATGATCC CATCACTTC TCAAAACCT
	1151	CATCAGACGT ATTGAAGATA AATAACGGCT CTGGGGAGGC TCTCAATCCA
30	1201	TATCAAGGAA CGATTCTATT TTCTGGAGAA ACCCTAACAG CAGATGAAC
	1251	TAAAGTTGCT GACAATTAA AATCTTCATT CACGCAGCCA GTCTCCCTAT
	1301	CCGGAGGAAA GTTATTGCTA CAAAAGGGAG TCACTTTAGA GAGCACGAGC
	1351	TTCTCTCAAG AGGCCGGTTC TCTCCTCGGC ATGGATTCAAG GAACGACATT
	1401	ATCAACTACA GCTGGGAGTA TTACAATCAC GAACCTAGGA ATCAATGTTG
	1451	ACTCTTCTAGG TCTTAAGCAG CCCGTCAGCC TAACAGCAA AGGTGCTCA
35	1501	ATAAAAGTGA TCGTATCTGG GAAGCTCAAC CTGATTGATA TTGAAGGGAA
	1551	CATTATGAA AGTCATATGT TCAGCCATGA CCAGCTCTC TCTCTATTAA
	1601	AAATCACGGT TGATGCTGAT GTTGATACTA ACGTTGACAT CAGCAGCCTT
	1651	ATCCCTGTTCTGCTGAGGA TCCTAATTCA GAATACGGAT TCCAAGGACA
	1701	ATGGAATGTT AATTGGACTA CGGATACAGC TACAAATACA AAAGAGGCCA
40	1751	CGGCAACTTG GACCAAAACA GGATTGTTCC CCAGCCCCGA AAGAAAATCT
	1801	GCGTTAGTAT GCAATACCCCT ATGGGGAGTC TTACTGACA TTGCTCTCT
	1851	GCAACAGCTT GTAGAGATCG GCGCAACTGG TATGGAACAC AAACAAGGTT
	1901	TCTGGGTTTC CTCCATGACG AACTTCCTGC ATAAGACTGG AGATGAAAAT
	1951	CGCAAAGGCT TCCGTACATC CTCTGGAGGC TACGTACATG GTGGAAAGTGC
45	2001	TCACACTCCT AAAGACGACC TATTACCTT TGCGTTCTGC CATCTCTTG
	2051	CTAGAGACAA AGATTGTTT ATCGCTACAA ACAACTCTAG AACCTACGGT
	2101	GGAACCTTAT TCTTCAGCA CTCTCATACC CTACAACCCC AAAACTATTT
	2151	GAGATTAGGA AGAGCAAAGT TTTCTGAATC AGCTATAGAA AAATCCCTA
50	2201	GGGAAATTCC CCTAGCTTGC GATGTCCAAG TTGCTTCAG CCATTCAGAC
	2251	AACCGTATGG AAACGCACTA TACCTCATTG CCAGAACCG AAGGTTCTTG
	2301	GAGCAACGAG TGTATAGCTG GTGGTATCGG CCTAGACCTT CCTTTGTTTC
	2351	TTCCAACCC ACATCCTCTT TTCAAGACCT TCATTCCACA GATGAAAGTC
	2401	GAAATGGTTT ATGTATCACA AAATAGCTTC TTCGAAAGCT CTAGTGATGG
	2451	CCGTGGTTT AGTATTGGAA GGCTGTTAA CCTCTCGATT CCTGTGGGTG
55	2501	CGAAATTCGT GCAGGGGAT ATCGGAGATT CCTACACCTA TGATCTCTCA
	2551	GGATTCTTTG TTTCCGATGT CTATCGTAAC AATCCCCAAT CTACAGCGAC
	2601	TCTTGTGATG AGCCCCAGACT CTTGGAAAAT TCGCGGTGGC AATCTTCAA
	2651	GACAGGCATT TTTACTGAGG GGTAGCAACA ACTACGTCTA CAACTCCAAT
	2701	TGTGAGCTCT CGGGACATTA CGCTATGGAA CTCCGTGGAT CTTCAAGGAA
60	2751	CTACAATGTA GATGTTGGTA CCAAACCTCG ATTCTAG

The PSORT algorithm predicts an outer membrane location (0.924).

The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 58A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 58B) and for FACS (Figure 58C) analyses. A GST-fusion protein was also expressed.

The cp6733 protein was also identified in the 2D-PAGE experiment (Cpn0451).

- 5 These experiments show that cp6733 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 59

The following *C.pneumoniae* protein (PID 4376814) was expressed <SEQ ID 117; cp6814>:

10	1 MHDALLSILA IQELDIKMR LMRVKKEHQK ELAKVQSLKS DIRRKVQEKE 51 LEMENLKTQI RDGENRIQEI SEQINKLENO QAAVKKMDEF NALTQEMITTA 101 NKERRSLEHQ LSDLMDKQAG GEDLIVSLKE SLASTENSS VIEKEIFESI 151 KKINEEGKAL LEQRTELKHA TNPELLSIYE RLLNNKKDRV VVPIENRVC 201 GCHIVLTPQH ENLVRKKDRL IFCEHCSRIL YWQESQVNAQ ENSTAKRRRR 251 RAAV*
----	--

- 15 The cp6814 nucleotide sequence <SEQ ID 118> is:

20	1 ATGCATGACG CACTTCTAAG CATTTGGCT ATTCAAGAGC TTGATATTAA 51 AATGATTCGC CTTATGCGCG TAAAGAAAGA ACATCAGAAA GAATTGGCTA 101 AAGTCCAATC TTTAAAAAGT GATATTCGTA GAAAAGTTCA GGAAAAAGAA 151 CTCGAAATGG AGAATTTGAA AACTCAAATT CGAGATGGAG AGAATCGCAT 201 CCAAGAGAGT TCTGAACAAA TCAATAAATT AGAAAATCAG CAAGCTGCTG 251 TAAAAAAAAT GGATGAGTTT AACGCTCTTA CCCAAGAAAT GACTACAGCA 301 AACAAAAGAAC GTCGCTCTTT AGAGCACCAG CTTAGCGATC TCATGGATAA 351 GCAAGCTGGA GGCAGAACCC TTATTGTCTC TCTAAAAGAA AGCTTAGCTT 401 CTACAGAAAA TAGTAGCAGT GTCATTGAAA AAGAAATTTC TGAAAGCATE 451 AAAAGAGATTA ATGAAGAAGG CAAAGCTTTC CTTGAACAAAC GGACAGAGTT 501 AAAGCATGCG ACGAATCCCG AACTACTCAG CATCTATGAG CGTCTATTAA 551 ACAATAAAAAA AGATCGCGTT GTGTTCTCTA TTGAAAATCG TGTCTGCAGT 601 GTTGTCTATA TTGTTCTAAC TCCTCAACAC GAAAATCTTG TAAGAAAGAA 651 AGACCCGACTC ATTTTTGCG AACATTGCTC TCGAATTCTC TATTGGCAAG 701 AATCCCAAGT CAATGCTCAG GAAAATTCCA CAGCAAAACG TCGTCGTCGT 751 CGCGCAGCTG TATAA
----	--

The PSORT algorithm predicts an inner membrane location (0.070).

- The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 59A) or his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used in Western blot (Figure 59B) and FACS (Figure 59C) analyses.

These experiments show that cp6814 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 60

The following *C.pneumoniae* protein (PID 4376830) was expressed <SEQ ID 119; cp6830>:

40	1 MKWLFPATAVF AAVLPALTAF GDPASVEIST SHTGSGDPTS DAALTGFTQS 51 STETDGTTYT IVGDTFFSTF TNIPPVVTP DANDSSSNSS KGGSSSSGAT 101 SLIRCSSNLHS DFDFTKDSVL DLYHLFFPSA SNTLNPAILLS SSSSGGSSSS 151 SSSSSSGSAS AVVAADPKGG AAFYSNEANG TLTFTTDSGN PGSLTLQNLK 201 MTGDGAIYS KGPLVFTGLK NLTFITGNESQ KSGGAAYTEG ALTTQAIVEA 251 VTFTGNTSAG QGGAIYVKEA TLFNALDSLK FEKNTSGQAG GGIYTESTLT 301 ISNITKSIEF ISNKASVPAP APEPTSPAPS SLINSTTIDT STLQTRAASA 351 TPAVAPVAAV TPPTPISTQET AGNGGAIYAK QGISISTFKD LTFKSNSASV
----	--

851 GFFVSDVYRN NPQSTATLVM SPD SWKIRGG NLSRQAFLLR GSNNYVYNSN
 901 CELFGHYAME LRGSSRNYNV DVG TKLRF*

A predicted signal peptide is highlighted.

The cp6733 nucleotide sequence <SEQ ID 116> is:

5	1	ATGAAGACTT	CGATTCCCTTG	GGTTTTAGTT	TCCTCCGTGT	TAGCTTTCTC
	51	ATGTCACCTA	CAGTCACTAG	CTAACGAGGA	ACTTTTATCA	CCTGATGATA
10	101	GCTTTAATGG	AAATATCGAT	TCAGGAACGT	TTACTCAGAA	AACTTCAGCC
	151	ACAAACATATT	CTCTAACAGG	AGATGTCTTC	TTTACGAGC	CTGGAAAAGG
	201	CACTCCCTTA	TCTGACAGTT	GTTTTAAGCA	AACCACGGAC	AATCTTACCT
10	251	TCTTGGGGAA	CGGTCACTAGC	TTAACGTTTG	GCTTTATAGA	TGCTGGCACT
	301	CATGCCAGGTG	CTGCTGCATC	TACAACAGCA	AATAAGAAC	TTACCTTCTC
	351	AGGGTTTTCC	TTACTGAGTT	TTGATTCCCTC	TCCTAGCACA	ACGGTTACTA
	401	CAGGTCAGGG	AACGCTTCC	TCAGCAGGAG	GCGTAAATT	AGAAAATATT
15	451	CGTAAACCTG	TAGTTGCTGG	GAATTTTCT	ACTGCAGATG	GTGGAGCTAT
	501	CAAAGGAGCG	TCTTCCCTT	TAACTGGCAC	TTCTGGAGAT	GCTCTTTTA
	551	GTAACAACTC	TTCATCAACA	AAGGGAGGAG	CAATTGCTAC	TACAGCAGGC
	601	GCTCGCATAG	CAAATAACAC	AGGTTATGTT	AGATTCCAT	CTAACATAGC
	651	GTCTACGTCA	GGAGGCGCTA	TCGATGATGA	AGGCACGTCG	ATACTATCGA
20	701	ACAACAAATT	TCTATATTTT	GAAGGGAATG	CAGCGAAAAC	TACTGGCGGT
	751	GCGATCTGCA	ACACCAAGGC	GAGTGGATCT	CCTGAACGTG	TAATCTCTAA
	801	CAATAAGACT	CTGATCTTG	CTTCAAACGT	AGCAGAAACA	AGCGGTGGCG
	851	CCATCCATGC	TAAAAAGCTA	GCCCTTCCCT	CTGGAGGCTT	TACAGAGTTT
	901	CTACGAAATA	ATGTCTCATC	AGCAACTCCT	AAGGGGGGTG	CTATCAGCAT
	951	CGATGCCTCA	GGAGAGCTCA	GTCTTCTG	AGAGACAGGA	AACATTACCT
25	1001	TTGTAAGAAA	TACCCTTACA	ACAACCGGAA	GTACCGATAC	TCCTAAACGT
	1051	AATGCGATCA	ACATAGGAAG	TAACGGGAAA	TTACACGGAAT	TACGGGCTGC
	1101	TAAAAATCAT	ACAATTCT	TCTATGATCC	CATCACTTCA	GAAGGAACCT
	1151	CATCAGACGT	ATTGAAAGATA	AATAACGGCT	CTGCGGGAGC	TCTCAATCCA
	1201	TATCAAGGAA	CGATTCTATT	TTCTGGAGAA	ACCCCTAACAG	CAGATGAAC
30	1251	TAAAGTTGCT	GACAATTAA	AATCTTCATT	CACCGAGCCA	GTCTCCCTAT
	1301	CCGGAGGAAA	GTATTGCTA	CAAAAGGGAG	TCACCTTAA	GAGCACGAGC
	1351	TTCTCTCAAG	AGGCCGGTTC	TCTCCTCGGC	ATGGATTTCAG	GAACGACATT
	1401	ATCAACTACA	GCTGGGAGTA	TTACAATCAC	GAACCTAGGA	ATCAATGTTG
35	1451	ACTCCTTAGG	TCTTAAGCAG	CCCAGTCAGCC	TAACAGAAA	AGGTGCTTCA
	1501	ATAAAAGTGA	TCGTATCTGG	GAAGCTAAC	CTGATTGATA	TTGAAGGGAA
	1551	CATTATGAA	AGTCATATGT	TCAGCCATGA	CCAGCTCTC	TCTCTATTAA
	1601	AAATCACGGT	TGATGCTGAT	GTGATACTA	ACGTTGACAT	CAGCAGCCTT
	1651	ATCCCTGTT	CTGCTGAGGA	TCTTAATTCA	GAATACGGAT	TCCAAGGACA
40	1701	ATGGAATGTT	AATTGGACTA	CGGATACAGC	TACAAATACA	AAAGAGGCCA
	1751	CGGCAACTTG	GACCAAAACA	GGATTGTTTC	CCAGCCCCGA	AAGAAAATCT
	1801	GCGTTAGTAT	GCAATACCT	ATGGGGAGTC	TTTACTGACA	TTCGCTCTCT
	1851	GCAACAGCTT	GTAGAGATCG	GCGCAACTGG	TATGGAACAC	AAACAAGGTT
	1901	TCTGGGTTTC	CTCCATGACG	AACTTCCTGC	ATAAGACTGG	AGATGAAAAT
45	1951	CGCAAAGGCT	TCCGTACATAC	CTCTGGAGGC	TACGTACATG	GTGGAAAGTGC
	2001	TCACACTCCT	AAAGACGACC	TATTTACCTT	TCCGTTCTGC	CATCTCTTG
	2051	CTAGAGACAA	AGATTGTTT	ATCGCTACCA	ACAACCTCTAG	AACCTACGGT
	2101	GGAACCTTAT	TCTTCAAGCA	CTCTCATACC	CTACAACCCC	AAAACATATT
	2151	GAGATTAGGA	AGAGCAAAGT	TTTCTGAATC	AGCTATAGAA	AAATCCCTA
50	2201	GGGAAATTCC	CCTAGCCTTG	GATGTCCAAG	TTTCGTTCTAG	CCATTCAGAC
	2251	AACCGTATGG	AAACGCACTA	TACCTCATTT	CCAGAACCCG	AAGGTTCTTG
	2301	GAGCAACGAG	TGTATAGCTG	GTGGTATCGG	CCTAGACCTT	CCTTTGTTTC
	2351	TTTCCAACCC	ACATCCTCTT	TTCAAGACCT	TCATTCCACA	GATGAAAGTC
	2401	GAAATGGTTT	ATGTATCAC	AAATAGCTTC	TTCGAAAGCT	CTAGTGATGG
55	2451	CCGTGGTTTT	AGTATTGGAA	GGCTGCTTAA	CCTCTCGATT	CCTGTGGGTG
	2501	CGAAATTCTGT	GCAGGGGGAT	ATCGGAGATT	CCTACACCTA	TGATCTCTCA
	2551	GGATTCTTTG	TTTCCGATGT	CTATCGTAAC	AATCCCCAAT	CTACAGCGAC
	2601	TCTTGTGATG	AGCCCAGACT	CTTGGAAAAT	TCGCGGTGGC	AATCTTCAA
	2651	GACAGGCATT	TTTACTGAGG	GGTAGCAACA	ACTACGTCTA	CAACTCCAAT
	2701	TGTGAGCTCT	TCGGACATTA	CGCTATGGAA	CTCCGTGGAT	CTTCAAGGAA
60	2751	CTACAATGTG	GATGTTGGTA	CCAAACTCCG	ATTCTAG	

The PSORT algorithm predicts an outer membrane location (0.924).

	1801	TACGTTACTA AAACCTTCCA GTGTTCCGAT TCTCATGCC TCCAGTTAC
5	1851	TAGTAATAAA GCAGCAGATG AAGGCGGGG CCTGTATTGT GGTGACGATG
	1901	TCACGCTAAC GAACCTGACA GGGAAAACAC TATTTCAAGA GAATAGCAGT
	1951	GAGAAACATG GAGGTGGCT CTCTCTGCC TCAGGAAAAT CTCTGACTAT
	2001	GACATCGTTA GAGAGCTTCT GCTTAAATGC AAATACAGCA AAGGAAAACG
	2051	GAGGCGGTGC GAATGTCCCT GAAAATATTG TACTCACCTT CACCTATACT
	2101	CCCACTCCAA ATGAACCTGC GCCTGTGCAG CAGCCCGTGT ATGGAGAAC
10	2151	TCTTGTACT GGAAATACAG CCACAAAAAG TGGTGGGGC ATTTACACGA
	2201	AAAATGCGGC CTTCTCAAT TTATCTCTG TAACTTTGTA TCAAAATACC
	2251	TCTTCAGAAA ATGGTGGTGC CTTACTTACC CAAAAAGCTG CAGATAAAAAC
	2301	GGACTGTTCT TTICACCTATA TTACAAATGT CAATATCAC C ACAATACAG
	2351	CTACAGGAAA TGGTGGGGC ATTGCTGGGG GAAAAGCACA TTTCGATCGC
15	2401	ATTGATAATC TTACAGTCCA AAGCAACCA GCAAAGAAAG GTGGTGGGGT
	2451	TTATCTTGAA GATGCCCTCA TCCGGAAA GGTATTACA GGTTCTGTCT
	2501	CACAAAATAC AGCTACAGAA AGTGGTGGGG GTATCTACGC TAAGGATATT
	2551	CAACTACAAG CTCTACCTGG AAGCTTCACA ATTACCGATA ATAAAGTCGA
	2601	AACTAGTCTT ACTACTAGCA CTAATTATA TGGTGGGGC ATCTATTCCA
	2651	GTGGAGCTGT CACGCTAAC AATATATCTG GAACCTTGG CATTACAGGA
20	2701	AACTCTGTIA TCAATACAGC GACATCCCAG GATGCAGATA TACAAGGTGG
	2751	GGGCATTAT GCAACCCAGT CTCTCTCAAT AAATCAATGT AATACACCCA
	2801	TTCTATTTAG CAACAACTCT GCTGCCACTA AAAAACATC AACAAACAAAG
	2851	CAAATTGCTG GTGGGGCTAT CTTCTCCGCT GCAGTAACTA TCGAGAATAA
	2901	CTCTCAGCCC ATTATTTCT TAAATAATT CGCAAAGTCG GAAGCAACTA
25	2951	CAGCAGCAAC TGCAGGAAAT AAAGATAGCT GTGGAGGAGC CATTGCAGCT
	3001	AACTCTGTIA CTTTAACAAA TAACCCCTGAA ATAACCTTA AAGGAAATTA
	3051	TGCAGAAACT GGAGGAGCGA TTGGCTGTAT TGATCTTAAT AATGGCTCAC
	3101	CTCCCCGTAA AGTCTCTATT GCAGACAACG GTTCTGTCT TTTCAAGAC
	3151	AACTCTGCGT TAAATCCGG AGGGCCTATC TATGGAGAGA CTATCGATAT
30	3201	CTCCAGGACA GGTGCGACTT TCATCGGTAA CTCTCAAAA CATGATGGAA
	3251	GTGCAATTTC CTGTTCAACA GCCCTAACCTC TTGCGCCAAA CTCCCAACTT
	3301	ATCTTGAAA ACAATAAGGT TACGGAAACC ACAGCCACTA CAAAAGCTTC
	3351	CATAAAATAAT TTAGGAGCTG CAATTATGG AAATAATGAG ACTAGTGACG
	3401	TCACTATCTC TTTATCAGCT GAGAATGGAA GTATTTCTT TAAAAACAAAT
35	3451	CTATGCACAG CAACAAACAA ATACTGCAGT ATTGCTGGAA ACGTAAAATT
	3501	TACAGCAATA GAAGCTTCAG CAGGGAAAGC TATATCTTC TATGATGCG
	3551	TTAACGTTTC CACCAAAGAA ACAATGCTC AAGAGCTAAA ATAAATGAA
	3601	AAAGCGACAA GTACAGGAAC GATTCTATT TCTGGGGAAC TTCACGAAAA
	3651	TAATCCTAT ATTCCACAGA AAGTCACTTT CGCACATGGG AATCTCATTC
40	3701	TAGGTAAAAA TGCAGAACTT AGCGTAGTTT CCTTTACCCA ATCTCCAGGC
	3751	ACCACAATCA CTATGGGCC AGGATCGGTT CTTTCCAACC ATAGCAAAGA
	3801	AGCAGGAGGA ATCGCTATAA ACAATGTCAT CATTGATTT AGTGAATCG
	3851	TTCTCTACTAA AGATAATGCA ACAGTAGCTC CACCCACTCT TAAATTAGTA
	3901	TCGAGAACTA ATGCAGATAG TAAAGATAAG ATTGATATTA CAGGAACGT
45	3951	GACTCTCTA GATCCTAATG GCAACTTATA TCAAAATTCT TATCTTGGTG
	4001	AAGACCGCGA TATCACTCTT TTCAATATAG ACAATTCTGC AAGTGGGGCA
	4051	GTACAGGCC CGAATGTCAC CCTCAAGGG AATTAGGAG CTAAAAAAGG
	4101	ATATTAGGA ACCTGGAATT TGGATCCAAA TTCTCGGGT TCAAAATTA
	4151	TTCTAAAATG GACCTTGAC AAATACCTGC GCTGGCCCTA CATCCCTAGA
50	4201	GACAACCACT TCTACATCAA CTCTATTGG GGAGCACAAA ACTCTTAGT
	4251	GACTGTGAAA CAAGGGATCT TAGGGAACAT GTTGAACAAAT GCAAGGTTG
	4301	AAGATCCTGC TTTCAACAAAC TTCTGGGCTT CGGCTATAGG ATCTTCCCT
	4351	AGGAAAGAAG TATCTCGAAA TTCTGACTCA TTACACTATC ATGGCAGAGG
	4401	CTATACCGCT GCTGTGGATG CCAAACCTCG CCAAGAATT TTCTTAGGAG
55	4451	CTGCCCTTCAG TCAGGTTTTT GGTCACGCC AGTCTGAATA TCACCTTGAC
	4501	AACTATAAGC ATAAAGGCTC AGGTCACTCT ACACAAGCAT CTCTTTATGC
	4551	TGGCAATATTC TTCTATTTC CTGCGATACG GTCTCGGCCT ATTCTATTCC
	4601	AAGGTGTGGC GACCTATGGT TATATGCAAC ATGACACCCAC AACCTACTAT
	4651	CCTTCTATTG AAGAAAAAAA TATGGCAAC TGGGATAGCA TTGCTGGTT
60	4701	ATTGATCTG CGTTTCAGTG TGGATCTAA AGAACCTCAA CCTCACTCTA
	4751	CAGCAAGGCT TACCTTCTAT ACAGAAGCTG AGTATACCAAG AATTGCCAG
	4801	GAGAAATTCA CAGAGCTAGA CTATGATCCT AGATCTTCT CTGCATGCTC
	4851	TTATGGAAAC TTAGCAATTCTC CTACTGGATT CTCTGTAGAC GGAGCATTAG
	4901	CTTGGCGTGA GATTATTCTA TATAATAAAAG TATCAGCTGC GTACCTCCCT
65	4951	GTGATTCTCA GGAATAATCC AAAAGCGACC TATGAAGTTC TCTCTACAAA
	5001	AGAAAAGGGC AACGTAGTCA ACGTTCTCCC TACAAGAAC GCAGCTCGTG
	5051	CAGAGGTGAG CTCTCAAATT TATCTTGAA GTTACTGGAC ACTCTACGGC
	5101	ACGTATACTA TTGATGCTTC AATGAATACT TTAGTGCAAA TGGCCAACCG
	5151	AGGGATCCGG TTTGTATTCT AG

401 DATLTVDSSST IGESGGAIFA ADSIQIQQCT GTTLFSGNTA NKSGGGIYAV
 451 GQVTLEDIAN LKMTNNTCKG EGGAIYTKKA LTINNGAILT TFSGNTSTDN
 501 GGAIFAVGGI TLSLDLVEVRF SKNKTGNNSA PITKAASNTA PVVSSSTTAA
 551 SPAVPAAAAA PVTNAAKGGA LYSTEGLTVS GITSLSFEN NECQNQGGGA
 601 YVTKTFQCSL SHRLQFTSNK AADEGGGLYC GDDVTLTNLT GKTLFQENSS
 651 EKHGGGLSLA SGKSLTMTSL ESFCLNANTA KENGGGANVP ENIVLTFTYT
 701 PTPNEPAPVQ QPVYGEALVT GNTATKSGGG IYTKNAAFSN LSSVTFDQNT
 751 SSENGGALLT QKAADKTDCS FTYITNVNIT NNTATGNGGG IAGGKAHFDR
 801 IDNLTVQSNO AKKGGGVYLE DALILEKVIT GSVSQNTATE SGGGIYAKDI
 851 QLQALPGSFT ITDNKVTDSL TTSTNLYGGG IYSSGAVTLT NISGTFGITG
 901 NSVINTATSQ DADIQGGGIY ATTSLSINQC NTPILFSNNS AATKKTSTTK
 951 QIAGGAIFSA AVTIENNSQP IIFLNNSAKS EATTAATAGN KDSCCGAIAA
 1001 NSVTLTNNPE ITFKGNYAET GGAIGCIDLT NGSPPRKVSI ADNGSVLFQD
 1051 NSALNRGGAI YGETIDISRT GATFIGNSSK HDGSAICCST ALTLAPNSQL
 1101 IFENNKVTTET TATTKASINN LGAAIYGNNE TSDVTISLSA ENGSIFFKNN
 1151 LCTATNKYCS IAGNVKFTAI EASAGKAISF YDAVNVTKE TNAQELKLNE
 1201 KATSTGTLIF SGELHENKSY IPQKVTFAHG NLILKGNAEL SVVSFTQSPG
 1251 TTITMGPGSV LSNHSKEAGG IAINNVIIDF SEIVPTKDNE TVAPP TLKLV
 1301 SRTNADSKDK IDITGTVTLL DPNGNLYQNS YLGEDRDITL FNIDNSASGA
 1351 VTATNVTLQG NLGAKKGYLG TWNLDPNSSG SKIILKWTFD KYLRWPYIPR
 1401 DHNFYINSIW GAQNSLVTVK QGILGNMLNN ARFEDPAFNN FWASAIGSFL
 1451 RKEVSRNSDS FTYHGRGYTA AVDAKPRQEF ILGAAFSQVF GHAESEYHLD
 1501 NYKHKGSGHS TQASLYAGNI FYFPAIRSRP ILFGQVATYG YMQHDTTYY
 1551 PSIEEKNMAN WDSIAWLFDL RFSDLKEPQ PHSTARLTFF TEAEYTRIRQ
 1601 EKFTELDYDP RSFSACSYGN LAIPTGFSD GALAWREIIIL YNKVSAAYLP
 1651 VILRNNPKAT YEVLSTKEKG NVVNVLPTRN AARAEVSSQI YLGSYWTLYG
 1701 TYTIDASMNT LVQMANGGIR FV*

A predicted signal peptide is highlighted.

The cp6830 nucleotide sequence <SEQ ID 120> is:

30 1 ATGAAGTGGC TACCAGCTAC AGCTGTTTT GCTGCCGTAC TCCCCGCACT
 51 51 AACAGCCTTC GGAGATCCCG CGTCTGTTGA AATAAGTACC AGCCATACAG
 101 101 GATCCGGGGA TCCTACAAGC GACGCTGCCT TAACAGGATT TACACAAAGT
 151 151 TCCACAGAAA CTGACGGTAC TACCTATACC ATTGTCCGGTG ATATCACCTT
 201 201 CTCTACTTTT ACGAATATTCT CTGTTCCCGT AGTAACCTCA GACCCAACG
 251 251 ATAGTTCCAG CAATAGCTCT AAAGGAGGAA GTAGCAGTAG TGGAGCTACA
 301 301 TCTCTAAATCC GATCCTCAAA CCTACACTCC GATTTTGATT TTACAAAAGA
 351 351 TAGCGTGTATC GACCTCTATC ACCCTTTCTT TCCCTCAGCT TCAAATACTC
 401 401 TCAATCCTGC ACTCCTTTCT TCCAGTAGCA GCGGTGGATC CTCGAGCAGC
 451 451 ATAGCTCCT CATCATCTGG AAGTGCATCT GCTGTTGTTG CTGGGGACCC
 501 501 AAAAGGAGGC GCTGCCTTT ATAGTAACGA GGCTAACCGGA ACTTTAACCT
 551 551 TCACTACAGA CTCTGGAAAT CCCGGCTCCC TGACTCTTCA GAATCTAAA
 601 601 ATGACCGGAG ATGGAGCCGC CATCTACTCG AAGGGTCCCT TAGTATTAC
 651 651 TGGTTTAAAAA AATCTAACCT TTACAGGAAA TGAATCTCG AAATCTGGAG
 701 701 GTGCTGCCCTA TACTGAAGGC GCACTCACAA CACAAGCAAT CGTTGAAGCC
 751 751 GTAACTTTTA CTGGCAACAC CTGGCAGGG CAAGGAGGCG CTATCTATGT
 801 801 TAAAGAAGCT ACCCTATTCA ATGCTCTAGA CAGCCTAAA TTTGAAAAAAA
 851 851 ACACCTCTGG GCAAGCTGGT GGTGGAATCT ATACAGAGTC TACGCTCACA
 901 901 ATCTCGAACAA TCACAAAATC TATTGAATT TCTCTAAATA AAGCTCTGT
 951 951 CCCTGCCCTC GCTCCTGAGC CCACCTCTCC GGCTCCAAGT AGCTTAATAA
 1001 1001 ATTCTACAAC GATCGATACC TCGACTCTCC AAACCCGAGC AGCATCCGCA
 1051 1051 ACTCCAGCAG TGGCTCTGT TGCTGCCGTAA CTCCAACAC CAATCTCTAC
 1101 1101 TCAAGAGACC GCAGGAAATG GAGGCCTAT CTATGCTAAA CAAGGTATTT
 1151 1151 CGATATCCAC GTTTAAAGAT CTGACCTTCA AGTCTAACCTC TGCATCGGT
 1201 1201 GATGCCACCC TTACTGTCGA TTCTAGCACT ATTGGAGAAT CTGGAGGTGC
 1251 1251 TATCTTGCA GCAGACTCTA TACAAATCCA ACAGTGCACG GGAACCACCT
 1301 1301 TATTCACTGG CAATACTGCC AATAAGTCTG GTGGGGGTAT TTACGCTGT
 1351 1351 GGACAAGTCA CCCTAGAAGA TATAGCGAAT CTGAAGATGA CCAACAACAC
 1401 1401 CTGTAAAGGT GAAGGTGGAG CCATCTACAC TAAAAAGGCT TTAACATATCA
 1451 1451 ACAACGGTGC CATTCTCACT ACATTTCTG GAAATACATC GACAGATAAT
 1501 1501 GGTGGGGCTA TTTTGTGT AGGTGGCATC ACTCTCTCTG ATCTTGTAGA
 1551 1551 AGTCCGCTTT AGTAAAAATA AGACCGGAAA TTATTCCGCT CCTATTACCA
 1601 1601 AAGCGGCTAG CAACACAGCT CCTGTAGTTT CTAGCTCTAC AACTGCTGCA
 1651 1651 TCTCCTGCGG TCCCTGCTGC CGCTGCAGCA CCTGTTACAA ACGCAGCAAA
 1701 1701 AGGAGGGGCT TTATATAGTA CAGAAGGACT GACTGTATCT GGAATCACAT
 1751 1751 CGATATTGTC GTTTGAAAAC AACGAATGCC AGAATCAAGG AGGTGGGGCT

Example 62

The following *C.pneumoniae* protein (PID 4377101) was expressed <SEQ ID 123; cp7101>:

5	1	MYSCYSKGIS HNYLLHPMSR LDIFVFDLSI ANQDQNLLEE IFCSEDTVLF
	51	KAYRTTALQS PLAAKNLNIA RKVANYILAD NGEIDTVKLV EAIHHLSQCT
	101	YPLGPHRHNE AQDREHLLKM LKALKENPKL KESIKTLFVP SYSTIQNLIR
	151	HTLALNPQTI LSTIHVROAA LTALFTYLRO DVGSCFATAP AILIHQEYPE
	201	RFLKDLDNDLI SSGKLSRIVN QREJAVPINL SGCIGELFKP LRILDLYPDP
	251	LVKLSSSPGL KKAFSAANLI ETLGDSEAQI QQLLSHQYLM QKLQNVHETL
10	301	TANDIIKSTL LHYYQLQUEST VRAIFFKEGL FSKEQVAFST QHPRELSEIQ
	351	RVYHYLHAYE EAKSAFIHDT QNPLLKAWEY TLATLADASQ PTISNHIRLA
	401	LGWKSEDPHS LVSLVTHFVE EEVENIRILV QCCEQTYHEA RSQLEYIEGR
	451	MRNPLNNQDS QILTMHDHMRF RQELNKALYE WDSAQEKAKK FLHLPEFLLS
	501	FYTQKQIPLYF RSSYDAFIQE FAHLYANAPA GFRILFTHGR THPNNTWSPIY
15	551	SINEFIRFLS EFFTSTESEL LGKHAVINLE KETSRLVHN1 TAMLHT'DVFQ
	601	EALLTRILEA YQLPVPPSIL NHLDQLSQTP WVVVSGGTVD TLLLDDYFESS
	651	EPLTLTEKHP ENPHELAAFY ADAALKDLPTG IKSYLEEGSH SLLSSSPTHV
	701	FSIIAGSPLF REAWDNDWYS YTWL RDVWVK QHQDFLQDTI LPQLSIYAFI
	751	ENFCNKYALQ HVVHDFHDFC SDHSLTLPFL YDKGSRFLSS LFTKDVTVAL
20	801	IYIRRLLYLM VREVVPVSEQ QLPEVLDNVS SYLGIISSRIT YEKFRLSLIEE
	851	TIPKMTLLSS ADLRHITYKGL LMQSYQK1YT EEDTYLRITT AMRHNNLAYP
	901	APLLFAFDSNW PSIYFGFILN PGTEIDLWK FNYAGLQQQP LDNIQELFAT
	951	SRPWTLYANP IDYGMPPPG YRSRLPKEFF *

The cp7101 nucleotide sequence <SEQ ID 124> is:

25	1	ATGTATTCTGT GTTACAGCAA AGGAATATCC CATAACTATC TTCTACATCC
	51	TATGTCACGT TTGGATATT TTGTTTCGA TTCTCTGATC GCAAACCAGG
	101	ATCAAAATCT TCTTGAGGAA ATTTCTGTT CTGAAGACAC AGTTTTATTT
	151	AAAGCCTACC GTACTACGGC TCTACAATCC CCTCTAGCTG CTAAGAACCT
	201	AAATATCGCC CGTAAAGTCG CAAATTATAT CTTAGCTGAC AATGGGGAAA
	251	TCGATACAGT AAAGCTTGTC GAAGCCATTC ACCATCTCTC ACAATGTACC
30	301	TATCCTTTAG GGCCTCATCG CCATAATGAA GCTCAAGATC GTGAACACCT
	351	CCTTAAAATG CTAAAAGCTC TAAAGGAAAA TCCTAAATTA AAAGAAAGCA
	401	TCAAAACTCT CTITGTCCTC TCATACTCTA CAATCCAAA CCTAATTCCG
	451	CATACACTAG CATTGAATCC ACAGACAATT CTCTCTACGA TTCATGTGCG
35	501	TCAAGCAGCA CTCACAGCGC TCTTCACCTA CCTTCGGCAA GATGTAGGTT
	551	CCTGTTTGC TACGGCTCCT GCCATTCTCA TTCACCAAGA ATATCCAGAA
	601	CGATTCCTTA AAGATCTAA TGATCTATT AGCAGTGGCA AACTCTCTAG
	651	AATCGTAAC CAAAGGGAAA TTGCGGTTCC TATAAACCTT TCGGGATGCA
	701	TTGGAGAGCT ATTCAAGCCT TTAAGGATTG TAGATCTTTA TCCTGATCCT
40	751	CTGGTTAACG TCTCCTCATC TCCAGGACTC AAAAAGCCT TTTCTGCTGC
	801	CAATCTTATT GAAACTCTTG GGGATTCTGA AGCACAAATC CAACAGTTGC
	851	TCTCGCATCA ATATTGATG CAAAAACTAC AAAATGTCCA TGAGACCTTA
	901	ACTGCTAACG ACATTATCAA ATCGACACCTT CTGCACTACT ATCAGCTCCA
	951	AGAAAGTACT GTACGAGCTA TTTCTTCAA AGAAGGGTG TTCAGCAAAG
45	1001	AACAAGTGGC ATTCTCGACG CAACACCCCCA GAGAGCTCTC AGAAATACAA
	1051	CGGGTATACC ACTACTACA TGCCTATGAA GAAGCAAAAT CTGCTTTAT
	1101	CCATGACACT CAAAATCCCT TACTGAAAGC CTGGGAGTAT ACTTAGCGA
	1151	CTCTTGGGAA TGCTAGCCAA CCTACCATCT CAAACCATAT CCGCCTTGCC
	1201	TTAGGATGGA AAAGTGAAGA CCCTCACAGT CTTGTATCTC TAGTTACACA
50	1251	CTTGTGAA GAGGAAGTAG AAAACATCCG AATTCTAGTC CAACAATGTG
	1301	AACAGACCTA TCACGAAGCA CGCTCCCAAC TAGAATATAT TGAAGGGCGG
	1351	ATGCGCAACC CACTAAATAA TCAAGACAGT CAGATTGTA CGATGGATCA
	1401	CATGCGCTTC CGTCAAGAAC TCAATAAAGC TCTTATGAG TGGGATAGTG
	1451	CTCAAGAAAA GGAAAGAAA TTTCTACATC TTCTGAATT CTTACTTCT
55	1501	TTCTATACAA AGCAATTCC CTTATACATT CGTAGTTCTT ACGATGCCCT
	1551	CATTCAAGAA TTTGCTCATC TCTATGCTAA TGCTCCCGCT GGCTTCCGTA
	1601	TTCTTTTCAc GCATGGACGC ACCCATCCGA ACACATGGTC CCCCATCTAT
	1651	TCGATTAATG AATTATACG TTTCTTTCT GAATTCTCTA CCTCCACAGA
	1701	GTCAGAACTT CTGGGAAAC ATGCCGTGAT CAATTAGAG AAAGAAACAT
	1751	CTCGGCTCGT CCACAAACATC ACTGCCATGC TACACACGGA TGTTTCAA
60	1801	GAAGCTCTCC TTACAAGAAT TTAGAAGCC TATCAGCTTC CTGTCCTCC
	1851	CTCCATCTTA AACCACTTAG ATCAGCTGTC ACAAACCCCC TGGGTTATG
	1901	TTCTGGAGG AACAGTGGAC ACTCTCTTT TGGATTATTT TGAAAGCTCA
	1951	GAACCTCTGA CACTTACAGA AAAGCATCTC GAAAATCCTC ATGAGCTTGC
	2001	AGCTTTCTAC GCAGACGCC TTAAAGATCT CCCTACAGGA ATTAAAAGTT

The PSORT algorithm predicts an outer membrane location (0.926).

The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 60A) or his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used in Western blot (Figure 60B) and FACS (Figure 60C) analyses.

- 5 The cp6830 protein was also identified in the 2D-PAGE experiment (Cpn0540) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6830 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 61

- 10 The following *C.pneumoniae* protein (PID 4376854) was expressed <SEQ ID 121; cp6854>:

```

1  MSIAIAREQY AAIIDMHPKP SIAMFSSEQA RTSWEKRQAH PYLYRLLEII
51  WGVVKFLLGL IFFIPLGLFW VLQKICQNFI LLGAGGWIFR PICRDSNLLR
101 QAYAARLFSA SFQDHVSSVR RVCLQYDEVF IDGLELRLPN AKPDRWMLIS
151 NGNSDCLEYR TVLQEKEWDI FRIAEEESQSN ILIFNYPGVM KSQGNITRNN
201 VVKSYQACVRY LRLDEPAGPQ ARQIVAYGYS LGASVQAEAL SKEIADGSDS
251 VRWFVVKDRG ARSTGAVAKQ FIGSLGVWLA NLTHWNINSE KRSKDLHCP
301 LFTIYGKDSQG NLIGDGLFKK ETCFAAPFLD PKNLEECSGK KIPVAQTGLR
351 HDHILSDDVI KEVAGHQRH FDN*

```

The cp6854 nucleotide sequence <SEQ ID 122> is:

```

20   1  ATGTCAATAG CTATTGCAAG GGAACAATAC GCAGCTATAT TGGATATGCA
      51  TCCTAAACCT TCGATGCCA TGTTTCTTC GGAGCAGGCG AGAACTTCTT
      101 GGGAGAAACG ACAGGCTCAT CCTTACCTTT ATCGTCTTCT TGAGATCATA
      151 TGGGGTGTTC TGAAATTCT TCTCGGCTTA ATCTTCTTTA TTCCCTTGGG
      201 TCTTTCTGG GTCCCTTCAGA AGATATGTCA GAATTTTATT CTTCTTGGTG
      251 CAGGAGGGTG GATTTTTAGA CCCATATGCA GGGACTCTAA TTTATTGCGA
      301 CAAGCTTACG CCGCGCGTCT TTTCTCCGCT TCATTCCAAG ATCATGTCTC
      351 CTCTGTGCGA AGGGTTTGCT TACAGTATGA CGAGGTCTTT ATTGACGGAT
      401 TGGAGTTACG TCTTCCCAAT GCTAAGCCAG ATCGATGGAT GTTAATCTCC
      451 AATGGAAACT CCGATTGCTT AGAGTATAGG ACAGTGCTGC AAGGGGAAAA
      501 GGACTGGATA TTCCGTATTG CTGAAGAGTC TCAATCCAAC ATTTAATCT
      551 TCAATTACCC AGGAGTCATG AAGAGCCAAG GGAATATAAC AAGAAACAAT
      601 GTAGTCAAAT CTTATCAAGC ATGCGTACGC TATCTTAGAG ATGAACCCGC
      651 AGGACCTCAG GCGCGTCAAA TCGTTGCTTA TGGCTATTCT TTAGGAGCTA
      701 GTGTTCAAGC CGAACCATTA AGTAAAGAGA TCGCAGACGG AAGTGATAGC
      751 GTCCGTGTTAGT TTGTCGTAA AGATCGAGGA GCTCGCTCTA CAGGAGCCGT
      801 TGCTAAACAG TTTATTGGAA GTCTAGGAGT TTGGCTGGCG AATCTTACCC
      851 ATTGGAATAT TAATTCTGAA AAGAGAAGCA AGGACTTGCA TTGCCCAGAA
      901 CTCTTATTT ATGGCAAGGA TTCCCAAGGT AATCTTATCG GGGATGGATT
      951 GTTCAAAAAAA GAGACGTGCT TCGCAGCACC ATTTTTAGAT CCTAAAAACT
      1001 TGGAAAGAGTG TTCAGGGAAAG AAAATCCCTG TAGCTCAGAC CGGTCTAAGA
      1051 CACGATCATC TCCCTTCCGA TGATGTGATT AAAGAAGTTG CAGGTCTAT
      1101 TCAAAGACAT TTGATAATT A

```

The PSORT algorithm predicts an inner membrane location (0.461).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 61A.

- 45 The recombinant protein was used to immunise mice, whose sera were used in Western blot (Figure 61B) and FACS (Figure 61C) analyses. A his-tagged protein was also expressed.

These experiments show that cp6854 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

5 651 TTCTATAATT GTAGGAACCA TGGTAGACGT GTCATGGAGA AATACCGCAG
 701 TACAATGGAT CGGGGATCAG CTCCTCTGTTA TTGGGACTTT AGGAGGAAC
 751 ACTTCTGTTG CTAGTGCAT CTCAACAGAT GGCACGTGTA TTGTAGGAGG
 801 TTCTGAAAAT GCAGATTCTC AGACTCATGC CTATGCTTAT AAAAACGGTG
 851 TTATGAGCGA TATAGGGACC CTCGGAGGTT TTTATTCTTT AGCACATGCA
 901 GTATCTTCAG ATGGTTCTGT GATTGTAGGA GTATCCACGA ACTCTGAGCA
 951 TAGATATCAT GCATTCCAAT ATGCTGATGG ACAGATGGTA GATTITAGGAA
 1001 CTTTAGGAGG GCCTGAATCT TATGCTCAAG GTGTTGCTGG AGATGGAAAG
 1051 GTAATTGTGG GTAGAGCACA AGTACCATCT GGAGATTGGC ATGCGTTCT
 1101 ATGTCCTTC CAAGCTCCGA GCCCTGCTCC TGTCATGGG GGAAGCACTG
 1151 TCGTAACTAG CCAGAACCTCA CGTGGAAATGG TAGATATCAA TGCTACGTAC
 1201 TCCCTTTGA AAAATAGCCA ACAACAACTA CAAAGATTGC TTATCCAGCA
 1251 TAGTGCAAAAA GTTGAAGATG TATCCTCAGG AGCACCATCT TTTACAAGTG
 1301 TGAAGGGTGC GATCTCAAAA CAGAGCCCTG CAGTGCAAAA TGATGTACAG
 1351 AAAGGGACGT TTTTAAGTTA CGGTTCCCAA GTTCATGGAA ACGTGCAGAA
 1401 TCAGCAATTG CTCACAGGAG CTTTTATGGA CTGGAAACTC GCTTCAGCTC
 1451 CTAATGCGG CTTTAAAGTA GCTCTCCACT ATGGCTCTCA AGATGCTCTC
 1501 GTAGAACGTG CAGCTCTTCC TTACACAGAA CAAGGCTTAG GAAGCAGTGT
 20 1551 CTTGTCAGGT TTTGGAGGAC AAGTTCAAGG ACGCTATGAC TTTAATTAG
 1601 GAGAAACTGT TGTTCGCAA CCCTTTATGG GCATTCAAGT TCTCCACCTA
 1651 AGTAGAGAAG GGTATTCTGA GAAGAATGTT CGATTTCCTG TAAGCTATGA
 1701 TTCTGTAGCC TACTCACCAG CTACTAGCTT TATGGGTGCG CATGTATTTG
 1751 CCTCCCTAAG CCCTAAAATG AGTACAGCAG CAACTTTAGG TGTGGAGAGA
 25 1801 GATCTGAATT CACATATAGA TGAATTAAAG GGATCCGTCT CTGCTATGGG
 1851 AAACTTTGTG TTGGAAAATT CTACAGTGAG TGTTTTAAGA CCTTTTGCTT
 1901 CTCTTGCTAT GTACTATGAC GTAAGACAAC AGCAACTCGT GACGTTGTCA
 1951 GTAGTTATGA ATCAACAACC CTTAACAGGC ACACTAAGCT TAGTAAGCCA
 2001 AAGTAGCTAT AATCTTAGCT TCTAA

The PSORT algorithm predicts an inner membrane location (0.100).

30 The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 63A) or his-tagged product. The proteins were used to immunise mice, whose sera were used in Western blot (Figure 63B) and FACS (Figure 63C) analyses.

These experiments show that cp7107 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

35 Example 64

The following *C.pneumoniae* protein (PID 4376467) was expressed <SEQ ID 127; cp6467>:

1 MLRFFAVFIS TLWLITSGCS PSQSSKGIFV VNMKEMPRSL DPGKTRLIAD
 51 QTLMRHLYEG LVEEHQSQNGE IKPALAESYT ISEDGTRYTF KIKNIILWSNG
 40 101 DPLTAQDFVS SWKEILKEDA SSVLYAFLP IKNARAIFDD TESPENLGV
 151 ALDKRHLEIQ LETPCAHFLH FLTLPIFFPV HETLRNYSSTS FEEMPITCGA
 201 FRPVSLEKGL RLHILEKNPMY HNKSRVKLHK IIVQFISNAN TAAILFKHKK
 251 LDWQGPPWGE PIPPEISASL HQDDQLFSLP GASTTWLLFN IQKKPWNNAK
 301 LRKALSLAID KDMILTKVVYQ GLAEPTDHIL HPRLYPGTYP ERKRQNERIL
 351 EAQQLFEEAL DELQMTREDL EKETLTFSFT SFSYGRICQM LREQWKKVLK
 45 401 FTIPIVGQEF FTIQKNFLEG NYSLTVNQWT AAFIDPMSYL MIFANPGGIS
 451 PYHLQDSHFQ TLLIKITQEH KKHLRNQLI EALDYLEHCH ILEPLCHPNL
 501 RIALNKNIKN FNLFVRRTSD FRFIEKL*

A predicted signal peptide is highlighted.

The cp6467 nucleotide sequence <SEQ ID 128> is:

50 1 ATGCTCCGTT TCTTCGGTGT ATTATATCA ACTCTTTGGC TCATTACCTC
 51 AGGATGTTCC CCATCCCAAT CCTCTAAAGG AATTTTTGTG GTAAATATGA
 101 AGGAAATGCC ACGCTCCCTG GATCCTGGAA AAACTCGTCT CATTGCAGAC
 151 CAAACTCTAA TGCCTCATCT ATATGAAGGA CTCGTCGAAG AACATTCCCA
 201 AAATGGAGAG ATTAAACCAG CCCTTGCAGA AAGCTACACC ATCTCCGAAG
 251 ACGGGACTCG GTACACATT AAAATCAAAA ACATCCTTG GAGTAACGGA
 301 GACCCCTCTGA CAGCTCAAGA CTTGTCCTCC TCTTGGAAAGG AAATCCTAAA

5 2051 ATCTAGAAGA AGGATCCCAC TCTCTACTTA GCTCATCACC CACCCACGTT
 2101 TTCTCTATAA TCGCAGGATC TCCTTATTT CGGGAAAGCTT GGGATAATGA
 2151 TTGGTACAGC TATACTGGC TTCGTGATGT CTGGGTGAAA CAACACCAAG
 2201 ATTTCCCTCA AGATACTATA TTACCTCAGC TAAGTATCTA TGCTTCATA
 2251 GAGAATTTTT GTAAACAAATA TGCTTGCAA CATGTAGTTC ATGACTTC
 2301 TGATTTCTGC TCCGACCACT CCTTGACTCT TCCGGAGCTC TATGACAAAG
 2351 GATCGCGTTT TCTAAGCTCC TTATTCACCA AAGATAAGAC CGTAGCTCTT
 2401 ATCTATATAC GCGCTCTCT CTACCTTATG GTCCGTGAAG TCCCTTATGT
 2451 TTCAAGAACAA CAGCTTCAG AAGTCTTACA TAACGTCTCT TCATATCTCG
 10 2501 GGATTTCTC TCGTATTAC TATGAGAAAT TCCGCTCCCT GATAGAGGAA
 2551 ACCATCCCTA AAATGACCTT ACTCTCCTCA GCAGACCTGA GGCATATCTA
 2601 TAAAGGTCTC CTCATGCAA GTTATCAAAA GATCTACACC GAAGAAAGATA
 2651 CGTACCTCCG CCTCACCAACG GCAATGAGGC ATCATAATCT TGCCTATCCC
 2701 GCTCCTTGC TCTTGAGA CAGTAACTGG CCTTCTATTT ATTTGGATT
 2751 CATCCTAAAT CCAGGAACCA CAGAGATCGA TCTTGGAAA TTAACTATG
 2801 CAGGGCTGCA AGGACAGCCT CTTGACAATA TCCAGGAGCT GTTCGCAACG
 2851 TCAAGACCCCT GGACCCCTCA TGCAAATCT ATAGATTATG GCATGCCACC
 2901 GCCTCCAGGC TACCGCAGCC GCCTCCCTAA AGAATTTTC TAG

The PSORT algorithm predicts a cytoplasmic location (0.206).

20 The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 62A) or his-tagged product. The proteins were used to immunise mice, whose sera were used in Western blot (Figure 62B) and FACS (Figure 62C) analyses.

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

25 These experiments show that cp7101 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 63

The following *C.pneumoniae* protein (PID 4377107) was expressed <SEQ ID 125; cp7107>:

30 1 MSIVRNSALP LPCLSRSETF KKVRSHMKFM KVLTPWIYRK DLWVTAFLLT
 51 AIPGSFAHTL VDIAGEPRHA AQATGVSGDG KIVIGMKVPD DPFAITVGFO
 101 YIDGHLQPLE AVRPOCSVYP NGITPDGTVI VGTNYAIGMG SVAVKWVNKG
 151 VSELPMLPDT LDVSASAVERA ALPDTLQVLSVYK QVLSVYK QVLSVYK
 201 SLPDAMNACV NGISSLGDSII VGTMVDVSWR NTAVQWIGDQ LSVIGTLGGT
 251 TSVASAISTD GTVIVGGSEN ADSQTHAYAY KNGVMSDIGT LGGFYSLAHA
 301 VSSDGSVIVG VSTNSEHRYH AFQYADGQMV DLGTLGGPES YAQGVSGDGK
 351 VIVGRAQVPS GDWHAFLCPF QAPSPAPVHG GSTVVTTSQNP RGMVDINATY
 401 SSLKNSQQQL QRLLIQHSAK VESVSSGAPS FTSVKGAIISK QSPAVQNDVQ
 451 KGTFLSYRSQ VHGNVQNQQL LTGAFMDWKL ASAPKCGFKV ALHYGSQDAL
 501 VERAALPYTE QGLGSSVLSG FGGQVQGRYD FNLGETVVLQ PFMGIQVLHL
 551 SREGYSEKNV RFPVSYDSVA YSAATSFMGA HVFASLSPKM STAATLGVER
 601 DLNSHIDEFK GSVSAMGNFV LENSTVSLR PFASLAMYD VRQQQLVTL
 651 VVMNQQPLTG TLSLVSQSSY NLSF*

The cp7107 nucleotide sequence <SEQ ID 126> is:

45 1 ATGAGTATAG TCAGAAATTC TGCATTGCCA CTTCCGTGTT TAAGCAGATC
 51 CGAAACCTTT AAAAAAGTTA GGTGCGATAT GAAATTATG AAAGTCCTTA
 101 CTCATGGAT TTATCGAAA GATCTTGGG TAACAGCATT CTTACTGACA
 151 GCAATTCCAG GATCTTGTGC ACATACTCTT GTTGATATAG CAGGAGAAC
 201 TCGGCATGCT GCTCAAGCAA CAGGAGTTTC TGGAGATGGT AAAATTGTTA
 251 TAGGAATGAA AGTTCCGGAT GATCCTTTG CTATAACTGT AGGATTCAA
 301 TATATTGATG GGCATTGCA ACCCTTAGAG GCAGTACGTC CTCAATGCTC
 351 TGTATACCTT AATGGTATAA CCCCGGACGG AACGGTTATT GTGGGTACAA
 401 ACTATGCCAT CGGGATGGGT AGTGTGCTG TGAAATGGGT AAATGGCAAG
 451 GTTTCTGAAT TTCCCATGCT CCCTGACACC CTCGATTCTG TAGCATCGGC
 501 AGTTTCTGCA GATGGAAGAG TGATTGGAGG GAATGAAAT ATAAATCTTG
 551 GCGCTTCTGT TGCTGTAAA TGGGAGGACG ACGTGATTAC ACAACTTC
 601 TCTCTTCTG ATGCTATGAA TGCTTGTGTT AACGGAATT CTTCAGATGG

601 GAAGTTGTTG CCAGAGTTGA GGGCTATGTT TGTGCTAACT ACTCGTAG

The PSORT algorithm predicts an inner membrane location (0.149).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 65A) and as a GST-fusion product (Figure 65B). The recombinant protein was used to immunise mice, whose sera were
5 used in a Western blot (Figure 65C) and for FACS analysis.

These experiments show that cp6679 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 66

The following *C.pneumoniae* protein (PID 4376890) was expressed <SEQ ID 131; cp6890>:

10	1 MKQOLLFCCVCFAMSCSAYAS PRRQDPSVMK ETFRNNYGI	VSGQEWKRG
	51 SDGTITKVLIK NGATLHEVYS GGLLHGEITL TFPHTTALDV VQIYDQGRLV	
	101 SRKTFVNGI PSQEELFNE D GTFVLTRWP NNDSDTITKP YFIETTYQGH	
	151 VIEGSYTSFVN GKYSSSIHNG EGVRSVFSSN NILLSEETFN EGVMVKYTTF	
15	201 YPNRDPEST HYQNQPHGL RLTYLQGGIP NTIEEWRYGF QDGTTIVFKN	
	251 GCKTSEIAYV KGVKEGLELR YNEQEIVAAE VSWRNDFLHG ERKIYAGGIQ	
	301 KHEWYYRGRS VSKAKFERLN AAG*	

A predicted signal peptide is highlighted.

The ep6890 nucleotide sequence <SEQ ID 132> is:

20	1 ATGAAACAAT TACTTTCTG TGTTTGGCTA TTTGCTATGT CATGTCTGC	
	51 TTACGCATCC CCACGACGAC AAGATCCTTC TGTTATGAAG GAAACATTCC	
	101 GAAATAATTA TGGCATTATT GTTTCCGGTC AAGAATGGGT AAAGCGTGGT	
	151 TCTGACGGCA CCATCACCAA AGTACTCAAA AATGGAGCTA CCCTGCATGA	
	201 AGTTTATTCT GGAGGCCCTCC TTCATGGGGA AATTACCTTA ACCTTTC	CCCC
25	251 ATACACACAGC ATTGGACGTT GTTCAAATCT ATGATCAAGG TAGACTCGTT	
	301 TCTCGCAAAA CCTTTTTTGT GAACGGTCTT CCATCTCAAG AAGAGCTGTT	
	351 CAATGAAGAT GGCACGTTTG TCCTCACACG ATGGCCGGAC AACAAACGACA	
	401 GTGATACCAT CACAAAGCCT TACTTCATAG AAACGACATA TCAAGGGCAT	
	451 GTCATAGAAC GAAGTTATAC TTCCCTTTAAT GGGAAATACT CCTCATCCAT	
30	501 CCACAATGGG GAGGGAGTTTC GTTCTGTGTT CTCTCCAAT AACATCCTTC	
	551 TTTCTGAAGA GACCTTCAT GAAGGTGTCA TGGTGAATA TACCACATTC	
	601 TATCCGAATC CGGATCCCAG ATCGATTACT CATTATCAAAT ATGGACAGCC	
	651 TCACGGCTTA CGGCTAACAT ATCTACAAGG TGGCATCCCC AATACGATAG	
	701 AGGAGTGGCG TTATGGCTTT CAAGACGGAA CGACCATCGT ATTTAAAAAT	
35	751 GGTGTAAAGA CATCTGAGAT CGCTTATGTT AAGGGAGTGA AAGAAGGTTT	
	801 AGAAACTGCGC TACAATGAAC AGGAAATTGT AGCTGAAGAA GTTCTTGCG	
	851 GTAATGATTT TCTGCATGGA GAACGTAAGA TCTATGCTGG AGGAATCCAA	
	901 AAGCATGAAT GGTATTACCG CGGGAGATCT GTATCTAAAG CCAAATTGCA	
	951 GCGGCTAAAT GCTGCAGGAT AG	

The PSORT algorithm predicts an outer membrane location (0.940).

40 The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 66A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 66B) and for FACS analysis. A his-tagged protein was also expressed.

These experiments show that cp6890 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 67

The following *C.pneumoniae* protein (PID 6172323) was expressed <SEQ ID 133; cp0018>:

5 351 GGAAGATGCG TCCTCCGTAT ATCTCTATGC GTTTTACCT ATCAAAATG
 401 CTCGGGCAAT CTTTGATGAT ACTGAGTCTC CAGAAAATCT AGGAGTCCGA
 451 GCTTAGATA AGCGTCATCT CGAAATTCTAG TTAGAAACTC CCTGCGCGCA
 501 TTCCCTACAT TTCTTGACTC TTCTTATTCTT TTTCCTGTT CATGAAACTC
 551 TGCGAAACTA TAGCACCTCT TTTGAAGAGA TGCCCATTAC CTGCGGTGCT
 601 TTCCGCCCTG TGTCTCTAGA AAAAGGCCCTG AGACTCCATC TAGAGAAAAA
 651 CCCTATGTAC CATAATAAAA GCCGTGTGAA ACTACATAAA ATTATTGTAC
 701 AGTTTATCTC AAACGCTAAC ACTGCAGCCA TTCTATTCAA ACATAAGAAA
 751 TTAGATTGGC AAGGACCTCC TTGGGGAGAA CCTATCCCTC CAGAAATCTC
 10 801 AGCTTCTCTA CATCAAGATG ACCAGCTCTT TTCTCTTCCG GGCGCTTCGA
 851 CTACATGGTT ACTCTTTAAT ATACAAAAAA AACCTTGAA CAATGCTAAA
 901 TTACGCAAGG CATTGAGCCT TGCAATAGAC AAAGATATGT TAACCAAAGT
 951 GGTATACCAA GGTCTTGAG AACCTACAGA TCATATCCTA CATCCAAGAC
 15 1001 TTATCCAGG GACCTATCCC GAACGGAAA GACAAAACGA AAGAATTCTT
 1051 GAGGCTCAAC AACTCTTGA AGAAGCTCTA GACGAACCTTC AAATGACACG
 1101 CGAAGATCTA GAAAAGGAAA CTTGACTTT CTCAACCTTT TCTTTTCTT
 1151 ACGGAAGGAT TTGCCAAATG CTAAGAGAAC AATGGAAGAA AGTCTTAAA
 1201 TTTACTATCC CTATAGTAGG CCAAGAGTTT TTCACAATAC AAAAAAAACTT
 1251 CCTAGAGGGG AACTATTCCC TAACCCTGAA CCAATGGACC GCAGCATTTA
 20 1301 TTGATCCGAT GTCTTATCTC ATGATCTTG CCAATCCTGG AGGAATTCTC
 1351 CCCTATCACC TCCAAGATTC ACACTTCAA ACTCTCTCA TAAAGATCAC
 1401 TCAAGAACAT AAAAAACACC TACGAAATCA GCTTATTATT GAAGCCCTTG
 1451 ACTATTTAGA ACACGTAC ATTCTCGAAC CACTATGTCA TCCAAATCTT
 1501 CGAATTGCTT TGAACAAAAA CATTAAAAAC TTTAATCTT TTGTTCGACG
 25 1551 AACCTCAGAC TTTCGTTTTA TAGAAAAACT ATAG

The PSORT algorithm predicts an outer membrane lipoprotein (0.790).

The protein was expressed in *E.coli* and purified as a his-tag product and a GST-fusion protein, as shown in Figure 64A. The recombinant his-tag protein was used to immunise mice, whose sera were used in a Western blot (Figure 64B). The recombinant GST-fusion protein was also used to immunise mice, whose sera were used in a Western blot (Figure 64C) and for FACS analysis (Figure 64D).

These experiments show that cp6467 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 65

35 The following *C.pneumoniae* protein (PID 4376679) was expressed <SEQ ID 129; cp6679>:

1 MRKMLVLLAS LGLLSPTLSS CTHLGSSGSY HPKLYTSGSK TKGVIAMLPV
 51 FHRPGKSLEP LPWNLQGEFT EEIISKRFYAS EKVFLIKHNA SPQTVSQFYA
 101 PIANRLPETI IEQFLPAEFI VATELLEQKT GKEAGVDSVT ASVRVRVF DI
 151 RHHKIALIYQ EIIIECSQPLT TLVNDYHRYG WNSKHFDSTP MGLMHSLRF
 40 201 EVVARVEGYV CANYS*

A predicted signal peptide is highlighted.

The cp6679 nucleotide sequence <SEQ ID 130> is:

45 1 ATGCGAAAAA TGTTGGTATT ATTGGCATCT TTAGGACTTC TATCCCCAAC
 51 CCTATCCAGC TGCACCTCACT TAGGCTCTTC AGGAAGTTAT CATCCTAACG
 101 TATACACTTC AGGGAGCAAA ACTAAAGGTG TGATTGCCAT GCTTCCCTGTA
 151 TTTCATCGCC CAGGAAAGAG TCTGAAACCT TTACCTTGGA ACCTCCAAGG
 201 AGAATTACT GAAGAGATCA GCAGAAAGTT TTATGCTTCG GAAAAGGTCT
 251 TCCTGATCAA GCACAATGCT TCACCTCAGA CAGTCTCTCA GTTCTATGCT
 301 CCGATTGCGA ATCGTCTACC CGAAACAATT ATTGAGCAAT TTCTCCCTGC
 351 AGAATTCACTT GTTGCTACAG AACTGTTAGA ACAAAAGACA GGGAAAGAAG
 401 CAGGTGTCGA TTCTGTAACA GCGCTGTAC GTGTTGCGGT TTTTGATATC
 451 CGTCATCATA AAATAGCTCT CATTATCAA GAGATTATCG AATGCAGCCA
 50 501 GCCTTTAACCT ACCCTAGTCA ATGATTATCA TCGCTATGGC TGGAACTCAA
 551 AACATTTTGA TTCAACGCCG ATGGGTTAA TGCAATAGCCG TCTTTCCGC

301 DDIDEESIRL QQAEAAALAR LPEEMSAFEG YIKVVESHLE NMKSLPYDGH
 351 GLEEKTQHQI RVVRSSLKAM VPEFLDIRRI FEEEEEFFLS ARKRLIDLAT
 401 TLVERKILTE QLERNNLRKA FSYLYQDSIF KKIIDNFEKL AWKFMILSKS
 451 ICRFTIIFEN HEHGVAKSLL HKNAVLLEKV IYRSLQKSYR DIGMSSAKMK
 501 ILHGNPFFSL EDNKKTIMKE HAEMLESLSS YRKVFLALSD ENVVDTPSDP
 551 KKWDLSGIPC RDALSEISRD EQWQKKAHLK HQESLYTQAR DRLTDQSSKE
 601 NQKELEKAEQ EYISSWERVK KFEIERVQER IRAIQKLYPN ILEREEETTG
 651 QETVTPTVQG TTASSDLTDI LGRIEVSSRE DNQNQESCVK VLRSHEVEMS
 701 WEVKQEYGPK KKEFQDQMGS LERFFTEHIE ELEVLFQKDYS KHLFSYFKVN
 751 NKKEVQYAKF RLKVLESDEL GILAQTESAE SLLTQEELPI LATRGALEKA
 801 VFKGSLCCAL ASKAKPYFEE DPRFQDSDTQ LRALTTLRQE AKASLEEEIK
 851 RFSNLENDIA EERRLLKESK QTFERAGLGV LREIAVESTY DLRSLTNTWE
 901 GTPESEKVYF SMYLNYYNEE KRRAKTRLVE MTQRYRDFKM ALEAMQFNEE
 951 ALLQEELSIQ APSE*

15 A predicted signal peptide is highlighted.

The cp6262 nucleotide sequence <SEQ ID 136> is:

1 ATGAGGAAAC TTCTGTATTCT TGCGATCGTT CTCATAGCTT TGAGCATTAT
 51 TTTGATTGCA GGTGGTGTGG TATTGCTTAC TGTAGCGATC CCTGGATTAA
 101 GTTCAGTCAT TTCTTCCCCG GCAGGGATGG GTGCCCTGTGC TTTGGGATGT
 151 GTGATGCTTG CTTTAGGGAT CGATGTTCTT CTGAAGAAC GAGAAGTCCC
 201 TATAGTTCTC GCATCTGTAA CTACGACACC AGGAACCTGGC AGCCCTAGAA
 251 GTGGTATTTTC TATTTCAGGA GCTGATAGCA CCATACGTC TCTTCCTACG
 301 TATCTCTTGG ACGAGGGACA TCCACAATCC ATGAGGAAAC TTCTGTATTCT
 351 TGGCATCGTT CTCATAGTTT TTAGCATTAT TTGATTGCA AGTGGTGTGG
 401 TATTGCTTAC TGTAGCGATC CCTGGATTAA GTTCAGTCAT TTCTTCCCCG
 451 GCAGGGATGG GTGCCCTGTGC TTTGGGATGT GTGATGCTTG CTTTAGGGAT
 501 CGATGTTCTT CTGAAGAAC GAGAAGTCCC TATAGTTCTC GCATCTGTAA
 551 CTACGACACC AGGAACCTGGC AGCCCTAGAA GTGGTATTTTC TATTTCAGGA
 601 GCTGATAGCA CCATACGTT TCTTCCTACG TATCCCTTGG ACGAGGGACA
 651 TCCACAATCC ATGAGGAAAC TTCTGTATTCT TGCGATCGTT CTCATAGTTT
 701 TTAGCATTAT TTGATTGCA AGTGGTGTGG TATTGCTTAC TGTAGCGATC
 751 CCTGGATTAA GCTCGATCAT TTCTTCCCCA GCGGAGATGG GTGCCCTGTGC
 801 TTGGGATGT GTGATGCTTG CTTGGGGAT CGACGTTCTT CTGAAGAAC
 851 GAGAAGTCCC TATAGTAGTT CCCGCACCTA TTCCCTGAAGA AGTCGTCATA
 901 GATGATATAG ATGAAGAGAG TATACGGCTG CAGCAGGAAG CTGAAGCCGC
 951 TTTAGCAAGA CTTCCCTGAG AGATGAGTGC ATTGAAAGGT TACATAAAAG
 1001 TTGTCGAGAG TCATTTGGAG AACATGAAAA GGCTGCCTTA TGATGGTCAT
 1051 GGCGCTAGAAG AGAAAACGAA ACATCAGATA AGAGTCGTC GATCTTCTTT
 1101 GAAGGCTATG GTTCCAGAAT TTTAGATAT CAGAAGAATT TTTGAAGAAG
 1151 AAGAGTTCTT TTTCTCTCA GCTCGCAAC GACTTATAGA TTTAGCTACT
 1201 ACTTTAGTAG AGAGAAAAAT TTTAACAGAG CAACTTGAGC GCAATAATTT
 1251 AAGGAAAGCG TTTTCTTATT TATATCAGGA CTCAATTTTT AAAAAAATTA
 1301 TTGATAACTT CGAGAAGTTA GCATGGAAAT TTATGATTTT GAGTAAATCA
 1351 ATTGTCGAT TTACAATTAT TTTGAAAAT CATGAACATG GTGAGCAA
 1401 GAGCCTGTTA CACAAGAATG CAGTGTACT GGAGAAGGTA ATCTATAGGA
 1451 GTTGTGCAAA AAGCTATAGA GATATAGGCA TGTGATCTGC AAAGATGAAA
 1501 ATCTTGCAAG GCAACCCCTT TTCTCTTTG GAAGATAATA AAAAGACGAT
 1551 AATGAAAGAA CACGCAGAGA TGCTTGAAG TCTCAGTAGC TATAGGAAGG
 1601 TATTTTTAGC TCTATCTGAT GAGAACGTT TAGATACACC TAGCGATCCA
 1651 AAGAAATGGG ATTTGTCAGG AATCCCTGT AGGGACGCGT TGTCTGAGAT
 1701 TTCTCGTGT GAAACAGTGGC AGAAGAAAGC ACATCTAAAG CATCAAGAGT
 1751 CCCTCTATAC GCAAGCTAGG GATCGTTAA CAGACCAGAG CTCTAAAGAA
 1801 AATCAGAAAG AGTTAGAGAA AGCTGAACAA GAGTACATAT CTTCTTGGGA
 1851 ACGGGTTAAA AAATTGAGA TTGAGAGAGT ACAGGAGAGG ATACGGGCAA
 1901 TTCAAAAGCT TTATCCTAAT ATCCTCGAGA GAGAAGAAGA AACACAGGT
 1951 CAGGAGACTG TGACTCCAAC TGTTCAAGGG ACGACGGCTT CATCCGATTT
 2001 AACAGATATT TTAGGAAGAA TAGAGGTCTC CAGTAGGGAG GATAATCAGA
 2051 ATCAAGAGTC TTGTGTAAGA GTCTTAAGAA GTCATGAGGT AGAAATGAGC
 2101 TGGGAAGTCA AACAAAGAGTA TGGCCCTAAG AAAAAAGAAT TTCAGGATCA
 2151 AATGGGTTCT TTAGAGAGGT TTTTACAGA GCATATTGAA GAGTTAGAAG
 2201 TATTACAGAA GGACTACTCT AACACATTGT CTTATTAA AAGTAAAC
 2251 AATAAGAAAG AGGTTCAATA TGCGAAGTTT AGCTTGAAGG TTTAGAGTC
 2301 AGATTTAGAA GGGATTCTAG CTCAGACTGA GAGTGCAG AGTCTGTTAA
 2351 CTCAAGAAGA ACTTCCGATT CTTGCAACTC GGGGAGCCTT AGAGAAAGCT
 2401 GTTTCAAGAAG GGAGTCATG TTGCGCGCTA GCAAGCAAAG CAAACCCCTA

5 1 MKTSVSMILLA LLCSGASSIV LHAATTPLNP EDGFIGEGNT NTFSPKSTTD
 51 AAGTTYSLTG EVLYIDPGKG GSITGTCFVE TAGDLTFLGN GNTLKFLSVD
 101 AGANIAVAHV QGSKNLNSFTD FLSLVITESP KSAVTIGKGS LVSLGAVQLQ
 151 DINTLVLTSN ASVEDGGVIK GNSCLIQGIK NSAIFGQNTS SKKGGAIISTT
 201 QGLTIENNLG TLKFNEMKAV TSGGALDLGA ASTFTANHEL IFSQNKTSGN
 251 AANGGAINCS GDLTFTDNTS LLLQENSTMQ DGGALCSTGT ISITGSDSIN
 301 VIGNTSGQKG GAISAASLKI LGQQGGALFS NNVVTHATPL GGAIFINTGG
 351 SLQLFTQGGD IVFEGNQVTT TAPNATTKRN VIHLESTAKW TGLAASQGNA
 10 401 IYFYDPITTN DTGASDMLRI NEVSANQKLS GSIVFSGERL STAEEAIAENL
 451 TSRINQPVTL VEGSLVLKQG VTLITQGFSQ EPESTLLLTL GTSL*

A predicted signal peptide is highlighted.

The cp0018 nucleotide sequence <SEQ ID 134> is:

15 1 ATGAAGACTT CAGTTCTAT GTTGTTGGCC CTGCTTTGCT CGGGGGCTAG
 51 CTTCTATTGTA CTCCCATGCCG CAACCACTCC ACTAAATCCT GAAGATGGGT
 101 TTATTGGGGA GGGCAATACA AATACTTTTT CTCCGAAATC TACAACGGAT
 151 GCTGCAGGAA CTACCTACTC TCTCACAGGA GAGGTTCTGT ATATAGATCC
 201 GGGGAAAGGT GGTTCAATT A CAGGAACCTTG CTTTGTAGAA ACTGCTGGCG
 251 ATCTTACATT TTTAGGTAAT GGAAATACCC TAAAGTTCCCT GTCGGTAGAT
 301 GCAGGGTCTA ATATCGCGGT TGCTCATGTA CAAGGAAGTA AGAATTAAAG
 351 CTTCACAGAT TTCCCTTCTC TGGTGATCAC AGAACATCCA AAATCCGCTG
 401 TTACTACAGG AAAAGGTAGC CTAGTCAGTT TAGGTGCAGT CCAACTGCAA
 451 GATATAAAACA CTCTAGTTCT TACAAAGCAAT GCCTCTGTCG AAGATGGTGG
 501 CGTGATTAAA GGAAACTCCT GCTTGATTCA GGGAAATCAA AATAGTGCGA
 551 TTTTGGACAA AAATACATCT TCGAAAAAAAG GAGGGGCGAT CTCCACGACT
 601 CAAGGACTTA CCATAGAGAA TAACTTAGGG ACGCTAAAGT TCAATGAAAA
 651 CAAAGCAGTG ACCTCAGGAG GCGCCTTAGA TTTAGGAGCC GCGTCTACAT
 701 TCACTGCGAA CCATGAGTTG ATATTTTCAC AAAATAAGAC TTCTGGAAAT
 751 GCTGCAAATG GCGGAGCCAT AAATTGCTCA GGGGACCTTA CATTACTGA
 801 TAACACTTCT TTGTTACTTC AAGAAAATAG CACAATGCAG GATGGTGGAG
 851 CTTTGTGTAG CACAGGAACC ATAAGCATTA CCGGTAGTGA TTCTATCAAT
 901 GTGATAGGAA ATACTPTCAGG ACAAAAAGGA GGAGCGATT CTGCAGCTTC
 951 TCTCAAGATT TTGGGAGGGC AGGGAGGCAG TCTCTTTCT AATAACGTAG
 1001 TGACTCATGC CACCCCTCTA GGAGGTGCCA TTTTTATCAA CACAGGAGGA
 1051 TCCTTGCAGC TCTTCACTCA AGGAGGGGAT ATCGTATTTCG AGGGGAATCA
 1101 GGTCACTACA ACAGCTCCAA ATGCTACCAC TAAGAGAAA GTAATTCCACC
 1151 TCGAGAGCAG CGCGAAGTGG ACGGGACTTG CTGCAAGTCA AGGTAACGCT
 1201 ATCTATTCT ATGATCCCAT TACCAAC GATAACGGAG CAACCGATAAA
 1251 CTACGTATC AATGAGGTC A GTGCAAATCA AAAGCTCTCG GGATCTATAG
 1301 TATTTTCTGG AGAGAGATTG TCGACAGCAG AAAGCTATAGC TGAAAATCTT
 1351 ACTTCGAGGA TCAACCAGCC TGTCACTTTA GTAGAGGGGA GCTTAGTACT
 1401 TAAACAGGGG GTGACCTTGA TCACACAAGG ATTCTCCAG GAGCCAGAAT
 1451 CCACGCTTCT TTTGGATCTG GGGACCTCAT TATAA

The PSORT algorithm predicts outer membrane (0.935).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 67A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 45 67B) and for FACS analysis.

These experiments show that cp0018 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 68

50 The following *C.pneumoniae* protein (PID 4376262) was expressed <SEQ ID 135; cp6262>:

55 1 MRKLRLILAIV LIALSIIILIA GGVVLLTVAI PGLSSVISSP AGMGACALGC
 51 VMLALGIDVL LKKREVPIVL ASVTTTPGTG SPRSGISISG ADSTIRSLPT
 101 YLLDEGHPQS MRKLRLILAIV LIVFSIILIA SGVVLLTVAI PGLSSVISSP
 151 AGMGACALGC VMLALGIDVL LKKREVPIVL ASVTTTPGTG SPRSGISISG
 201 ADSTIRSLPT YPLDEGHPQS MRKLRLILAIV LIVFSIILIA SGVVLLTVAI
 251 PGLSSIISSP AEMGACALGC VMLALGIDVL LKKREVPIVV PAPIPEEVVI

-110-

1401 AAAGCTTGC TCTCTACGTC TTGATGAAAA AGAGTTATTA CAAAAAGAAA
 1451 TCAAGAAAGA GGAATTATC CAAAAGAAC AACAAAGGCA TGCAGATAGA
 1501 TCACGTACATA CTACGTATCA AAAGCTACGA ATTGCTGAAG AGCTTGCTCT
 1551 TGAGCTGAAG AAGAAAATCT AA

- 5 The PSORT algorithm predicts cytoplasmic location (0.412).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 69A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 69B) and for FACS analysis.

These experiments show that cp6269 is a surface-exposed and immunoaccessible protein, and that it
 10 is a useful immunogen. These properties are not evident from the sequence alone.

Example 70

The following *C.pneumoniae* protein (PID 4376270) was expressed <SEQ ID 139; cp6270>:

1 MKIPLRFLLI SLVPTLSMSN LLGAATTEEL SASNSFDGTT STTSFSSKTS
 5 SATDGTNYVF KDSVVIENVP KTGETQSTSC FKNDAAAGDLNFLGGGF SFT
 10 FSNIDATTAS GAAIGSEAAN KTVTLSGFSA LSFLKSPAST VTNGLGINV
 15 KGNLSLLDND KVLIQDNFST GDGGAINCAG SLKIANNKSL SFIGNSSSTR
 20 GGAIIHTKNLT LSSGGETLFO GNTAPTAAGK GGAIAIAIDSG TLSISGDSD
 25 IIFEGNTIGA TGTVSHSAID LGTSAKITAL RAAQGHTIYF YDPITVTGST
 30 SVADALNINS PDTGDNKEYT GTIVFSGEKL TEAEAKDEKN RTSKLLQNVA
 35 FKNGTVVLKG DVVLSANGFS QDANSKLIIMD LGTSLVANTE SIELTNLEIN
 40 IDSRLRNGKKI KLSAATAQKD IRIDRPVLA ISDESFYQNG FLNEDHSYDG
 45 ILELDAGKDI VISADSRSID AVQSPYGYQG KWTINWSTD KKATVSWAKQ
 50 SFNPTAEQEA PLVPNLLWGS FIDVRSFQNF IELGTEGAPY EKRFWVAGIS
 55 NVLHRSGREN QRKFRHVSGG AVVGASTRMP GGDTLSLGFA QLFARDKDYF
 60 MNTNFAKTYA GSLRLQHDAS LYSVVSILLG EGGLREILLP YVSKTLPCSF
 65 YGQLSYGHTD HRMKTESLPP PPPTLSTDHT SWGGYVWAGE LGTRVAVENT
 70 SGRRGFFQEYT PFVKVQAVYA RQDSFVELGA ISRDFSDSHL VNLAIPLGIK
 75 LEKRAEAEQYY HVVAMYSPDV CRSNPKCTTT LLSNQGSWKT KGSNLARQAG
 80 IVQASGFRSL GAAAELFGNF GFEWRGSSRS YNVDAGSKIK F*

- 30 A predicted signal peptide is highlighted.

The cp6270 nucleotide sequence <SEQ ID 140> is:

1 ATGAAGATTC CACTCCGCTT TTTATTGATA TCATTAGTAC CTACGCCCTTC
 5 TATGTCGAAT TTAAATTAGGAG CTGCTACTAC CGAAGAGTTA TCGGCTAGCA
 10 ATAGCTTCGA TGGAACTACA TCAACAAACAA GCTTTCTAG TAAAACATCA
 15 TCGGCTACAG ATGGCACCAA TTATGTTTT AAAGATTCTG TAGTTATAGA
 20 AAATGTACCC AAAACAGGGG AAACTCAGTC TACTAGTTGT TTTAAAAATG
 25 ACGCTGCAGC TGGAGATCTA AATTCTTAG GAGGGGGATT TTCTTTCACA
 30 TTTAGCAATA TCGATGCAAC CACGGCTCT GGAGCTGCTA TTGGAAGTGA
 35 AGCAGCTAAT AAGACAGTCA CGTTATCAGG ATTTCCGCA CTTTCTTTTC
 40 TAAATCCCC AGCAAGTACA GTGACTAATG GATTGGGAGC TATCAATGTT
 45 AAAGGGAAATT TAAGCCTATT GGATAATGAT AAGGTATTGA TTCAGGACAA
 50 TTTCTCAACA GGAGATGGCG GAGCAATTAA TTGTGCAGGC TCCTTGAAGA
 55 TCGCAAACAA TAAGTCCCTT TCTTTATTG GAAATAGTTC TTCAACACGT
 60 GCGGGAGCGA TTCATACCAA AACCTCACA CTATCTCTG GTGGGGAAAC
 65 TCTATTTCAG GGGAAATACAG CGCCTACGGC TGCTGGTAAA GGAGGGTGC
 70 TCGCGATTGC AGACTCTGGC ACCCTATCCA TTTCTGGAGA CAGTGGCGAC
 75 ATTATCTTTG AAGGCAATAC GATAGGAGCT ACAGGAACCG TCTCTCATAG
 80 TGTATTGAT TTAGGAAC TCGCTAAAGAT AACTGCGTTA CGTGCCTGC
 85 AAGGACATAC GATATACTTT TATGATCCGA TTACTGTAAC AGGATCGACA
 90 TCTGTTGCTG ATGCTCTCAA TATTAATAGC CCTGATACTG GAGATAACAA
 95 AGAGTATACG GGAACCTAG TCTTTCTGG AGAGAAGCTC ACGGAGGGCAG
 100 AAGCTAAAGA TGAGAAGAAC CGCACTTCTA AATTACTTCA AAATGTTGCT
 105 TTAAAAATG GGACTGTAGT TTAAAGGTT GATGTCGTT TAAGTGC
 110 CGGTTCTCT CAGGATGCAA ACTCTAAGGT GATTATGGAT TTAGGGACGT
 115 CGTTGGTTGC AACACCGAA AGTATCGAGT TAACGAATT GGAAATTAAT
 120 ATAGACTCTC TCAGGAACGG GAAAAAGATA AAACCTCAGTG CTGCCACAGC

5
2451 TTTTGAAGAG GATCCCAGAT TCCAAGATT TGATACGAA TTGCGAGCTC
2501 TGACTCTAAG GTTACAGGGAG GCTAAGGCAGA GCCTGGAAAGA AGAGATAAAG
2551 AGATTTCAA ATCTTGAGAA CGATAITGCA GAGGAAAGAC GCCTTCTTAA
2601 AGAGAGCAAG CAGACGTTG AAAGAGCAGG TTTAGGGTT CTCCGAGAAA
2651 TTGCACTCGA GTCTACTTAT GATTGCGTT CCTTAACAAA TACATGGAA
2701 GGGACCCCAAG AGAGTGAGAA GGTCTATTAA AGCATGTATC TTAATTATTA
2751 CAACGAAGAG AACCGTAGGG CTAAAACAAG ATTGGTTGAA ATGACACAGA
2801 GGTATAGAGA TTTTAAAATG GCCTTGGAG CTATGCAGTT TAATGAAGAA
2851 GCCCTTTGC AAGAGGAAC CTCTATTCAA GCTCCCAGTG AATAA

10 The PSORT algorithm predicts inner membrane (0.660).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 68A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 68B) and for FACS analysis.

15 These experiments show that cp6262 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 69

The following *C.pneumoniae* protein (PID 4376269) was expressed <SEQ ID 137; cp6269>:

20
1 MYQENLRLL RLLYNSVQKS YADRLFSYEK TKMVHDTPLI PWEEDKEKCA
51 EAEKAFLQQ KILLDYGKSI FWLNENDEIN LNDPWSWGLN TVRTRKVQFQE
101 VDDSERWNHK VLIQKLEDDY EKLLEESSKE STEANKKLLS DLVDRLED
151 TKFFLKKQEE VETRVKDLRA RYGGTVDPKQ DTEAKKKVEL EASLETFLDS
201 IESELVQCLE DQDIYWKEQD VKDLARTQEL EEQDIEAKRE EAAEDLRSLN
251 ERLKKSKTML DRAKWHIENA EDSITWWTSQ IEMKDMKARL KILKEDITSV
301 LPEIDEIETC LSLEELPLLT TRELLTKSYL KFKICSETLL KMTSVFENNI
351 YVQEYEVOLQ NLGFKLQGIS QRFGKKQDDF ANLEEQVALQ KKRLRELTQN
401 FEIQQFNFMK EDFKAAAKDL YIRSTAEQKM NFDVPCMELF RRYHEEVNKP
451 LLELMYNCAD SYRDAKKKLC SLRLDEKELL QKEIKKEEFY QKKQQRHADR
501 SRHTTYQKLR IAEELALELK KKI*

The cp6269 nucleotide sequence <SEQ ID 138> is:

30
1 ATGTACCAGG AGAACCTAAG ATTGTTGGAA AGGCTTCCTT ATAATAGTGT
51 TCAAAAGAGC TATGCGGATC GGCTGTTTC CTATGAAAAG ACAAAAGATGG
101 TGCACGATAC TCCGCTGATT CCTTGGGAAG AGGATAAGGA AAAATGTGCT
151 GAAGCTGAGA AAGCTTCTT AGAGCAACAG AAGATTCTCC TAGATTATGG
201 AAAATCTATC TTTTGGCTGA ATGAGAACGA TGAGATCAAT TTAAACGATC
251 CTTGGAGTTG GGGCTTAAT ACGGTGAGGA CTAGGAAAGT ATTCCAAGAG
301 GTTGACGACA GTGAACGTTG GAATCATTAAG GTACTCATTC AAAAACTCGA
351 GGACGATTAT GAGAAACCTTC TAGAGGAAAG TTCAAAAGAG TCTACTGAAG
401 CAAATAAGAA GCTTTATCT GACTTÀGTAG ATCGTCTTGA AGATGCTAAG
451 ACAAAATTTT TCCTGAAGAA ACAGGAGGAG GTGGAGACTC GCGTTAAGGA
501 TCTTAGAGCT CGATATGGAG GCACAGTAGA TCCTAACGAG GATACGGAAG
551 CTAAGAACAA AGTCGAATTG GAGGCTAGCT TAGAAACCTT TTTAGATTCC
601 ATCGAATCAG AGCTAGTACA GTGTTTAGAA GATCAAGATA TATATTGGAA
651 AGAACAGGAT GTCAAAGATC TAGCACGTAC GCAAGAGCTC GAGGAACAAG
701 ATATTGAAGC GAAGAGGGAA GAAGCTGCCG AAGACCTAAG AAGTCTTAAT
751 GAGCGTTAA AGAACGCTAA AACTATGTTA GATAGGGCTA AATGGCATAT
801 TGAAGATGCT GAGGACAGTA TTACCTGGTG GACTAGTCAG ATAGAAATGA
851 AGGATATGAA AGCAAGACTG AAGATCTTAA AAGAAGATAT AACAAAGTGT
901 CTACCTGAAA TAGATGAGAT TGAAACGTGT TTAAGCTTAG AGGAGCTTCC
951 TTGCTTACG ACCAGGAAC TCTTAACTAA GTCTTACCTA AAGTTAAGA
1001 TTGTTCCGA AACACTATTA AAAATGACTT CTGTTGTTGA GAACAATATC
1051 TATGTTCAAGG AGTACGAGGT TCAGCTGCAA AATCTAGGGT TTAAGTTACA
1101 AGGTATATCT CAGAGATTG AAAGAAAACA AGACGATTTT GCGAATCTAG
1151 AGGAACAGGT TGCTTGCAG AAGAAACGAC TCAGAGAGCT CACTCAGAAT
1201 TTGAAATAC AAGGATTCAA TTTCATGAAA GAAGATTAA AGGCAGCCGC
1251 TAAAGATCTT TATATAAGAA GTACAGCTGA ACAAAAGATG AACTTTGATG
1301 TGCTTGCAT GGAGCTCTC CGTAGGTATC ATGAGGGAGGT CAACAAAGCCG
1351 CTCTTGTAGT TGATGTACAA TTGTGCAGAC AGTTATAGAG ATGCTAAGAA

701 TAAAAAGCGA ATTTCTTATT TCCACAACCT TTATAGATAC GGCCAACCCC
 751 TTCTAA

The PSORT algorithm predicts cytoplasmic (0.158).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 71A). The

5 recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 71B) and for FACS analysis.

These experiments show that cp6402 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 72

10 The following *C.pneumoniae* protein (PID 4376520) was expressed <SEQ ID 143; cp6520>:

1 MKHYLSFPSPS ADFFSKQGAI ETQVLFGERV LVKGSTCYAY SQLFHINELLW
 51 KPYPGHSFRS TLVPCTPEFH IHPNVSVVSV DAFLDPWGIP LPFGTLLHVN
 101 SQNTVIFPKD ILNHMNTIWG SGTPQCDPRH LRRLNYNFNA ELLIKDADLL
 15 151 LNFPLYVWGGR SVHESLEKPG VDCSGFINIL YQAQGYNVPR NAADQYADCH
 201 WISSFENLPS GGLIFLYPK EKRISHVMLK QDSSTLIHAS GGGKKVEYFI
 251 LEQDGKFLLDS TYLFFRNNQR GRAFFGIPRK RKAFL*

The cp6520 nucleotide sequence <SEQ ID 144> is:

1 ATGAAACACT ACCTATCATT TTCTCCCTCT GCTGATTTTT TCTCTAAACA
 51 GGGTGCTATT GAAACTCAAG TCCTTTTGG AGAGCGCGTC TTAGTCAAAG
 101 GGAGCACCTG CTATGCATAT TCCCAATTAT TCCACAATGA GCTGTTATGG
 151 AAGCCCTATC CAGGTCTAG CTTTCGTTCT ACCCTAGTCC CCTGCACTCC
 201 TGAATTTCAT ATCCATCCAA ATGTTTCTGT GGTTTCTGTG GATGCATTTT
 251 TAGATCCTTG GGGGATCCCT CTTCCCTTTG GAACTTTACT CCATGTGAAT
 301 TCTCAAAATA CCGTTATTTT CCCTAAGGAT ATTCTCAATC ATATGAACAC
 351 CATCTGGGGC TCCGGCACAC CTCAATGCGA TCCTAGACAT CTACGTCGTC
 401 TAAATTATAA CTTCTTTGCT GAACTTTAA TTAAAGACGC AGACCTTTA
 451 CTGAACCTTC CCTATGTATG GGGAGGACGG TCTGTACACG AAAGTCTGGA
 501 AAAGCCGGGT GTTGATTGTT CGGGATTTAT CAATATCCTT TACCAGGCAC
 551 AGGGATACAA CGTCCCTAGA AACGCTGCAG ATCAATATGC GGATTGTCAT
 601 TGGATCTCTA GCTTTGAGAA CCTTCCTTCT GGTGGGTTAA TATTCTTTA
 651 CCCTAAAGAA GAAAAGCGTA TTTCTCATGT TATGTTGAAA CAGGATAGTT
 701 CCACCCCTCAT TCATGCTTCT GGTGGAGGGAA AAAAGTGGAA GTATTTCATT
 751 TTAGAACAAAG ATGGGAAGTT TTTAGATTG ACTTATCTAT TTTTAGAAA
 801 TAATCAGAGG GGACGGGCAT TTTTGGGAT CCCTAGAAA AGAAAAGCCT
 851 TTCTGTAA

The PSORT algorithm predicts cytoplasmic (0.265).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 72A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 72B) and for FACS analysis.

40 These experiments show that cp6520 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 73

The following *C.pneumoniae* protein (PID 4376567) was expressed <SEQ ID 145; cp6567>:

1 MTSPIPFQSS GDASFLAEQP QQLPSTSESQ LVTQLLTMMK HTQALSETVL
 51 QQQRDRRLPTA SIIILQVGGAP TGGAGAPFQP GPADDHHHPI PPPVVPAQIE
 101 TEITTIIRSEL QLMRSTLQQS TKGARTGVLV VTAILMTISL LAIIIIILAV
 151 LGFTGVLPQV ALLMQGETNL IWAMVSGSII CFIALIGTLG LILTMNKNTPL

1251 TCAGAAAGAT ATTCGTATAG ATCGCCTGT TGTACTGGCA ATTAGCGATG
 1301 AGAGTTTTA TCAAAATGGC TTTTGATGAG AGGACCATTC CTATGATGGG
 1351 ATTCTTGAGT TAGATGCTGG GAAAGACATC GTGATTTCTG CAGATTCTCG
 1401 CAGTATAGAT GCTGTACAAT CTCGTATGG CTATCAGGGA AAGTGGACGA
 1451 TCAATTGGTC TACTGTATGAT AAGAAAGCTA CGGTTTCTG GGCAGAACAG
 1501 AGTTTTAATC CCACTGCTGA GCAGGGAGGCT CCGTTAGTTC CTAATCTTCT
 1551 TTGGGGTTCT TTTATAGATG TTGCTTCCTT CCAGAATTTC ATAGAGCTAG
 1601 GTACTGAAGG TGCTCCCTAC GAAAAGAGAT TTTGGGTTGC AGGCATTTC
 1651 AATGTTTGC ATAGGAGCGG TCGTGAAAAT CAAAGGAAAT TCCGTCATGT
 1701 GAGTGGAGGT GCTGTAGTAG GTGCTAGCAC GAGGATGCGG GGTGGTGATA
 1751 CCTTGCTCTC GGCTTTGCT CAGCTCTTG CGGGTGACAA AGACTACTTT
 1801 ATGAATACCA ATTCGCAAA GACCTACGCA GGATCTTAC GTTTGCAGCA
 1851 CGATGCTTCC CTATACTCTG TGGTGAGTAT CCTTTAGGA GAGGGAGGAC
 1901 TCCGCCAGAT CCTGTTGCCT TATGTTTCCA AGACTCTGCC GTGCTCTTC
 1951 TATGGGCAGC TTAGCTACGG CCATACGGAT CATCGCATGA AGACCGAGTC
 2001 TCTACCCCCC CCCCCCCCAGA CGCTCTCGAC GGATCATACT TCTTGGGGAG
 2051 GATATGTCG GGCTGGAGAG CTGGGAACTC GAGTTGCTGT TAAAAATACC
 2101 AGCGGCAGAG GATTTTCCA AGAGTACACT CCATTTGTA AAGTCCAAGC
 2151 TGTTTACGCT CGCCAAAGATA GCTTTGTAGA ACTAGGAGCT ATCAGTCGTG
 2201 ATTTTAGTGA TTCGCATCTT TATAACCTTG CGATTCTCT TGGAAATCAAG
 2251 TTAGAGAAAC GGTTTGAGA GCAATATTAT CATGTTGTTAG CGATGTATTG
 2301 TCCAGATGTT TGTGTTAGTA ACCCCAAATG TACGACTACC CTACTTTCCA
 2351 ACCAAGGGAG TTGGAAGACC AAAGGTTCGA ACTTAGCAAG ACAGGCTGGT
 2401 ATTGTTCAAGG CCTCAGGTTT TCGATCTTG GGAGCTGCAG CAGAGCTTTT
 2451 CGGGAACTTT GGCTTTGAAT GGCGGGGATC TTCTCGTAGC TATAATGTAG
 2501 ATGCGGGTAG CAAAATCAA TTTAG

The PSORT algorithm predicts outer membrane (0.92).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 70A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot and for FACS analysis (Figure 70B).

The cp6270 protein was also identified in the 2D-PAGE experiment (Cpn0013).

These experiments show that cp6270 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 71

35 The following *C.pneumoniae* protein (PID 4376402) was expressed <SEQ ID 141; cp6402>:

1 MNVADLLSHL ETLLSSKIFQ DYGPNGLQVG DPQTPVKKIA VAVTADLETI
 51 KQAVAAEANV LIVHHGIFWK GMPPYPIITGMI HKRIQLLIEH NIQLIAYHLP
 101 LDAHPTLGNN WRVALDLNWH DLKPFGSSLP YLGVQGSFSP IDIDSFIDLL
 151 SOYYQAPLKG SALGGPSRVS SAALISGGAY RELSSAATSQ VDCFITGNFD
 201 EPAWSTALES NINFLAFGHT ATEKVGPKSL AEHLKSEFPI STTFIDTANP
 251 F*

The cp6402 nucleotide sequence <SEQ ID 142> is:

1 ATGAATGTTG CGGATCTCCT TTCTCATCTT GAGACTCTTC TCTCATCAA
 51 AATATTTCAAG GATTATGGAC CCAACGGACT TCAAGTTGGA GATCCCCAAA
 101 CTCGGTAAA GAAAATCGCT GTTGCAGTTA CCGCAGATCT AGAAACCATA
 151 AAACAAGCTG TTGCGGGCAGA AGCAAACGTT CTCATTGTTAC ACCACGGAAT
 201 TTTTTGGAAA GGTATGCCCT ATCCTATTAC CGGCATGATC CATAAGCGCA
 251 TCCAATTACT AATAGAACAC AATATCCAC TCATTGCCCTA CCACCTTCCT
 301 TTGGATGCTC ACCCTACCTT AGGAAATAAC TGGAGAGTTG CCCTGGATCT
 351 AAATTGGCAT GACTTGAAAGC CCTTGGTTTC TCCCTCCCT TATTTAGGAG
 401 TCGAAGGCTC TTCTCTCCCT ATCGATATAG ATTCTTTCAT TGACCTGTTA
 451 TCTCAATATT ACCAAGCTCC CCTAAAAGGA TCTGCGCTTGG GCGGGCCCTC
 501 TAGAGTCTCC TCAGCAGCTC TGATCTCAGG AGGAGCTTAT AGAGAACTCT
 551 CTCGGCAGC CACGTCCCAA GTCGATTGCT CTCATCACAGG AAATTTGAT
 601 GAAACCTGCAT GGTCGACAGC TCTAGAAAGC AATATCAACT TCCTAGCATT
 651 TGGACATACA GCCACAGAAA AAGTAGGTCC AAAATCTCTT GCAGAGCATC

5 651 TACAAGTTGG TTTACTGGAG CTGGACTCTA TCACCCAGAT ATTGTTGAAC
 701 AAGATAGCTT GGCAATTACG AATTACCTAC ATAATAACGG GTACGCTGAT
 751 GCTATAGTCA ACTCTCACTA TGACCTTGAC GACAAAGGGA ATATTCTTCT
 801 TTACATGGAT ATTGATCGAG GGTGCGATA TACCTTAGGA CACGTCCTATA
 851 TCCAAGGGTT TGAGGTTTTG CCAAAACGCC TTATAGAAAA GCAATCCCAA
 901 GTCGGCCCCA ATGATCTTTA TTGCCCCGAT AAAATATGGG ATGGGGCTCA
 951 TAAGATCAA CAAACTATG CAAAGTATGG CTACATCAAT ACCAATGTAG
 1001 ACGTTCTCTT CATCCCTCAC GCAACCCGCC CTATTTATGA TGTAACCTTAT
 1051 GAGGTAAGTG AAGGGTCTCC TTATAAAGTT GGGTTAATTAA AAATTACTGG
 1101 GAATACCCAT ACAAAATCTG ACGTTATTTT ACACGAAACC AGTCTCTTCC
 1151 CAGGAGATAC ATTCAATCGC TTAAAGCTAG AAGATACTGA GCAACGTTA
 1201 AGAAAATACAG GCTACTTCA AAGCGTTAGT GTCTATACAG TTCGTTCTCA
 1251 ACTTGATCCT ATGGGCAATG CGGATCAATA CCGAGATATT TTTGTAGAAG
 1301 TCAAAGAAC AACAAACAGGA AACTTAGGCT TATTCTTAGG ATTTAGTTCT
 1351 CTGACAATC TTTTGAGG AATTGAACTA TCTGAAAGTA ATTTTGATCT
 1401 ATTGGAGCT AGAAATATAT TTCTAAAGG TTTTCGTTGT CTAAGAGGCG
 1451 GTGGAGAAC A TCTATTCTTA AAAGCCAATC TCAGGGACAA AGTCACAGAC
 1501 TATACTTGA AGTGGACCAA ACCTCATTTT CTAAACACTC CTTGGATTTT
 1551 AGGAATTGAA TTAGATAAT CAATTAACAG AGCATTATCT AAAGATTATG
 1601 CTGTCCAAAC CTATGGGGGG AACGTCAGCA CAACGTATAT CTTGAACGAA
 1651 CACCTGAAAT ACGGTCTATT TTATCGAGGA AGTCAAACGA GTTTACATGA
 1701 AAAACGTAAG TTCCCTCTAG GGCCAAATAT AGACAGCAAT AAAGGATTTG
 1751 TCTCTGCTGC AGGTGTCAAC TTGAATTACG ATTCTGTAGA TAGTCCCTAGA
 1801 ACTCCAACTA CAGGGATTG CGGGGGGGTG ACTTTGAGG TTTCTGGTTT
 1851 GGGAGGAAC TATCATTAA CAAAACCTC TTAAACAGC TCTATCTATA
 1901 GAAAACCTAC GCGTAAAGGT ATTTTAAAAA TCAAAGGGGA AGCTCAATT
 1951 ATAAACCCCT ATAGCAATAC TACAGCTGAA GGAGTTCTG TCAGTGAGCG
 2001 CTTCTTCTA GGTGGAGAGA CTACAGTTCG GGGATATAAA TCCTTTATTA
 2051 TCGGTCCAAA ATACTCTGCT ACAGAACCTC AGGGAGGAAC CTCTTCGCTC
 2101 CTTATTTCAAGAGTTCA ATACCCCTCTC ATCAGACAAAC CTAATATTAG
 2151 TGCCTTTGTA TTCTTAGACT CAGGTTTTGT CGGTTTACAA GAGTATAAGA
 2201 TTTCGTTAAA AGATCTACGT AGTAGTGCTG GATTTGGTCT GCGCTTCGAT
 2251 GTAATGAATA ATGTTCTGT TATGTTAGGA TTTGGTTGGC CCTTCCGTCC
 2301 AACCGAGACT TTGAATGGAG AAAAAATTGA TGTATCTAG CGATTCTTCT
 2351 TTGCTTTAGG GGGCATGTTCA TAA

The PSORT algorithm predicts outer membrane (0.7658).

The protein was expressed in *E.coli* and purified as GST-fusion (Figure 74A), his-tag and his-tag/GST-fusion products. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 74B) and for FACS analysis (Figure 74C).

40 The cp6576 protein was also identified in the 2D-PAGE experiment (Cpn0300).

These experiments show that cp6576 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 75

The following *C.pneumoniae* protein (PID 4376607) was expressed <SEQ ID 149; cp6607>:

45 1 MNKRQDKLKLK ICVIISTLIL VGIFARAPRG DTFKTFLKSE EAIYSNQCN
 51 EDRMKILCDA IEHADEEIFL RIYNLSEPKI QQLSLTRQAQA KNKVTIYYQK
 101 FKIPQILKQA SNVTLVEQPP AGRKLMHQKA LSIDKKDAWL GSANYTNLSL
 151 RLDNNNLILGM HSSELCDLII TNTSGDFSIK DQTGKYFVLP QDRKIAIQAV
 201 LEKIOTAQKT IQVAMFALTH SEIIQALHQA KQRGIHVDDII IDRSHSKLTF
 251 KQLRQLNINK DFVSINTAPC TLHHKFAVID NKTLLAGSIN WSKGRFSLND
 301 ESLIILENLT KQQNQKLRMI WKDLAKHSEH PTVDDEEKEI IEKSLPVEEQ
 351 EAA*

A predicted signal peptide is highlighted.

The cp6607 nucleotide sequence <SEQ ID 150> is:

201 PAS*

The cp6567 nucleotide sequence <SEQ ID 146> is:

```

5      1 ATGACCTCAC CGATCCCC TTCAAGTCTAGT GGCGATGCCT CTTTCCTTGC
      51 CGAGCAGCCA CAGCAACTCC CGTCTACTTC TGAATCTCAG CTAGTAACTC
     101 AATTGCTAAC CATGATGAAG CATACTCAAG CATTATCCGA AACGGTTCTT
     151 CAACAACAAAC GCGATCGATT ACCAACCGCA TCTATTATCC TTCAAGTAGG
    201 AGGAGCTCCT ACAGGAGGAG CGGGTGCCTC TTTTCAACCA GGACCGGCAG
    251 ATGATCATCA TCATCCCATA CCGCCGCCCTG TTGTACCAGC TCAAATAGAA
    301 ACAGAAATCA CCACTATAAG ATCCGAGTTA CAGCTCATGC GATCTACTCT
    351 ACAACAAAGC ACAAAAGGAG CTCGTACAGG AGTTCTAGTG GTTACTGCAA
    401 TCTTAATGAC GATCTCCCTA TTGGCTATTAA TTATCATAAT ACTAGCTGTG
    451 CTGGATTAA CGGGCGCTCT GCCTCAAGTA GCCTTATTGA TGCAAGGGTGA
    501 AACAAATCTG ATTTGGGCTA TGGTGAGCGG TTCTATTATT TGCTTTATTG
    551 CGCTAATTGG AACTCTAGGA TTAATTAA CAAATAAGAA CACGCCCTCA
   601 CCGGCTTCTT AA

```

The PSORT algorithm predicts inner membrane (0.694).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 73A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 73B) and for FACS analysis.

20 These experiments show that cp6567 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 74

The following *C.pneumoniae* protein (PID 4376576) was expressed <SEQ ID 147; cp6576>:

```

25      1 MLIMRNKVIL QISILALIQT PLTLFSTEKV KEGHVVVDSI TIITEGENAS
      51 NKHPLPKLKT RSGALFSQLD FDEDLRILAK EYDSVEPKVE FSEGKTNIAL
     101 HLLAKPSIRN IHISGNQVVP EHKLKTLQI YRNDLFEREK FLKGQLDDLR
     151 YYLKRGYFAS SVDYSLEHNQ EKGHIDVLIK INEGPGKIK QLTFSGISRS
     201 EKSDIQEFIG TKQHSTTSW FTGAGLYHPD IVEQDSLAIT NYLHNNGYAD
     251 AIVNSHYDLD DKGNILLYMD IDRGSRVTLG HVHIQGFEVL PKRLIEKQSQ
     301 VGPNDLYCPD KIWDGAHKIK QTYAKGYIN TNVDVLFIPH ATRPIYDVY
     351 EVSEGSPYKV GLIKITGNTH TKSDVILHET SLFPGDTFNR LKLEDTEQRL
     401 RNTGYFQSVS VYTVRSQLDP MGNADQYRDI FVEVKETTG NLGLFLGFSS
     451 LDNLFGGIEL SESNFDFLFGA RNIFSKGFRC LRGGGEHLFL KANFGDKVTD
     501 YTLKWTKPHF LNTPWILGIE LDKSINRALS KDYAVQTYGG NVSTTYILNE
     551 HLKYGLFYRG SQTSLHEKRK FLLGPNIDSN KGFVSAAGVN LNYDSVDSPR
     601 TPTTGIRGGV TFEVSGLGGT YHFTKLSLNS SIYRKLTRKG ILKIKGEAQF
     651 IKPYSNTTAE GVPVSERFFL GGETTVRGYK SFIIGPKYSA TEPOQGLSSL
     701 LISEEFAQYPL IRQPNISAFV FLDSGFVGLO EYKISLKDLR SSAGFGLRFD
     751 VMNNVPVMLG FGWPFRPTET LNGEKIDVSQ RFFFALGGMF *

```

40 A predicted signal peptide is highlighted.

The cp6576 nucleotide sequence <SEQ ID 148> is:

```

45      1 ATGCTCATCA TCGAAATAA AGTTATCTTG CAAATATCTA TTCTAGCGTT
      51 AATCCAAACC CCTTTAACCTT TATTTTCTAC TGAAAAAGTT AAAGAAGGCC
     101 ATGTGGTGGT AGACTCTATC ACAATCATAA CGGAAGGAGA AAATGCTTCA
     151 AATAAACATC CCTTACCCAA ATTAAAGACC AGAAGTGGGG CTCTTTTTTC
     201 TCAATTAGAT TTTGATGAAG ACTTGAGAAT TCTAGCTAAA GAATACGACT
     251 CTGTTGAGCC TAAAGTAGAA TTTTCTGAAG GGAAACTAA CATAGCCCTT
     301 CACCTAATAG CTAAACCCCTC AATTCGAAAT ATTCAATATCT CAGGAAATCA
     351 AGTCGTTCCCT GAACATAAAA TTCTTAAAC CCTACAAATT TACCGTAATG
     401 ATCTCTTGA ACGAGAAAAA TTTCTTAAGG GTCTTGATGA TCTAAGAACG
     451 TATTATCTCA AGCGAGGATA TTTCGCATCC AGTGTAGACT ACAGTCTGGA
     501 ACACAATCAA GAAAAGGTC ACATCGATGT TTTAATTAAA ATCAATGAAG
     551 GTCCTTGCAG GAAAATTAAA CAGCTTACGT TCTCAGGAAT CTCTCGATCA
     601 GAAAATCAG ATATCCAAGA ATTATTCAA ACCAAGCAGC ACTCTACAC

```

951 ATTAGGAGGG GTGGCTCTTG AATGTCAAAG ATGA

The PSORT algorithm predicts inner membrane (0.168).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 76A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 76B) and for 5 FACS analysis.

The cp6624 protein was also identified in the 2D-PAGE experiment.

These experiments show that cp6624 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 77

10 The following *C.pneumoniae* protein (PID 4376728) was expressed <SEQ ID 153; cp6728>:

1	MKSSVSWLFF SSIPLFSSLS IVAAEVTLDS SNNSYDGNSNG TTFTVFSTTD
51	AAAGTTYSLL SDVSFQNAGA LGIPLASGCF LEAGGDLTFQ GNQHALKFQF
101	INAGSSAGTV ASTSAADKNL LFNFDSRLSI ISCPSSLSP TGQCALKSVD
151	NLSLTGNSQI IFTQNFSSDN GGVINTKNFL LSGTSQFASF SRNQAFTGKQ
201	GGVVYATGTI TIENSPGIVS FSQNLAKGSG GALYSTDNCs ITDNFQVIFD
251	GNSAWEAAQA QGGAICCTT DKTVTLTGNK NLSFTNNNTAL TYGGAISGLK
301	VSIISAGGPTL FQSNISGSSA GQGGGGAINI ASAGELALSA TSGDITFNMM
351	QVTNGSTSTR NAINIIDTAK VTSIRAATGQ SIYFYDPITN PGTAASTDTL
401	NLNLDANSE IEYGGAIIVS GEKLSPTKEA IAANVTSTIR QPAVLARGDL
451	VLRDGVTVTF KDLTQSPGSR ILMDGGTTLS AKEANLSSLNG LAVNLSSLDG
501	TNKAALKTEA ADKNISLSGT IALIDTEGSF YEHNLKSAS TYPILLELTTA
551	GANGTITLGA LSTLTLQEPE THYGYQGNWQ LSWANATSSK IGSINWTRTG
601	YIPSPERKSN LPLNSLWGNF IDIRSINQLI ETKSSGEFFE RELWLSGIAN
651	FFYRDSMPTR HGFRHISGGY ALGITATTTPA EDQLTFAFCQ LFARDRNHIT
701	GKNHGDTYGA SLYFHHTEGL FDIANFLWGK ATRAPWVLSE ISQIIPLSFD
751	AKFSYLHTDN HMKTYYTDNS IIKGSRNDA FCADLGASLP FVISVPYLLK
801	EVEPFVKVQY IYAHQQDFYE RHAEGRAFNK SELINVEIPI GVTFERDSKS
851	EKGTYDLTLM YILDAYRRNP KCQTSLIASD ANWMAYGTNL ARQGFSVRAA
901	NHFQVNPHME IFGQFAFEVR SSSRNYNTNL GSKFCF*

30 The cp6728 nucleotide sequence <SEQ ID 154> is:

1	ATGAAGTCCT CTGTCTCTTG GTTGTCTTT TCTTCAATCC CGCTCTTTTC
51	ATCGCTCTCT ATAGTCGGC CAGAGGTGAC CTTAGATAGC AGCAATAATA
101	GCTATGATGG ATCTAACCGA ACTACCTTC CCGTCTTTTC CACTACGGAC
151	GCTGCTGCAG GAACTACCTA TTCCCTACTT TCCGACGTAT CCTTCAAAA
201	TGCAGGGGCT TTAGGAATT CCTTAGCCTC AGGATGCTTC CTAGAACCGG
251	GCGCGATCT TACTTCCA GGAAATCAAC ATGCACTGAA GTTGCATTT
301	ATCAATGCGG GCTCTAGCGC TGGAACTGTA GCCAGTACCT CAGCAGCAGA
351	TAAGAATCTT CTCTTTAATG ATTTTCTAG ACTCTCTATT ATCTCTTGTC
401	CCTCTCTTCT TCTCTCTCCT ACTGGACAAT GTGCTTTAAA ATCTGTGGGG
451	AATCTATCTC TAACTGGCAA TTCCCAAATT ATATTTACTC AGAACCTCTC
501	GTCAGATAAC GGCGGTGTTA TCAATACGAA AAACCTCTTA TTATCAGGGA
551	CATCTCAGTT TCGCAGCTTT TCGAGAAACC AAGCCTTCAC AGGGAAAGCAA
601	GGCGGTGTTAG TTTACGCTAC AGGAACCTATA ACTATCGAGA ACAGCCCTGG
651	GATAGTTTCC TTCTCTCAA ACCTAGCGAA AGGATCTGGC GGTGCTCTGT
701	ACAGCACTGA CAACTGTTG ATTACAGATA ACTTTCAAGT GATCTTGAC
751	GGCAATAGTG CTTGGGAAGC CGCTCAAGCT CAGGGCGGGG CTATTGTTG
801	CACTACGACA GATAAAACAG TGACTCTTAC TGGGAACAAA AACCTCTCTT
851	TCACAAAATAA TACAGCATTC ACATATGGCG GAGGCCATCTC TGGACTCAAG
901	GTCAGTATTG CCGCTGGAGG TCCTACTCTA TTTCAAAGTA ATATCTCAGG
951	AAGTAGCGCC GGTCAAGGGAG GAGGAGGAGC GATCAATATA GCATCTGCTG
1001	GGGAACCTCGC TCTCTCTGCT ACTTCTGGAG ATATTACCTT CAATAACAAAC
1051	CAAGTCACCA ACGGAAGCAC AAGTACAAGA AACGCAATAA ATATCATTGA
1101	TACCGCTAAA GTCACATCGA TAGGAGCTGC TACGGGGCAA TCTATCTATT
1151	TCTATGATCC CATCACAAAT CCAGGAACCG CAGCTTCTAC CGACACATTG
1201	AACTTAAACT TAGCAGATGC GAACAGTGTAG ATCGAGTATG GGGGTGCGAT
1251	TGTCTTTCT GGAGAAAAGC TTTCCCCTAC AGAAAAAGCA ATCGCTGCAA

5

```

1 ATGAATAAAA GACAAAAAGA TAAATTAAAA ATCTGTGTTA TTATTAGCAC
51 GTTGATTTA GTAGGAATTT TTGCAAGAGC TCCTCGTGGT GACACTTTA
101 AGACTTTTTT AAAGTCTGAA GAAGCTATCA TCTACTCAA TCAATGCAAT
151 GAGGACATGC GTAAAATCT ATGCCATGCT ATAGAACACG CTGATGAAGA
201 GATCTTCCTA CGTATTIATA ACCTCTCAGA ACCCAAGATC CAACAGAGTT
251 TAACTCGACA AGCTCAAGCA AAAAACAAAG TTACGATCTA CTATCAAAAAA
301 TTAAATTC CCCAAATCTT AAACCAAGCC AGCAATGTA CTTTAGTCGA
351 GCAACCTCCA GCAGGGCGTA AACTGATGCA TCAAAAGCT CTTTCATAG
401 ATAAGAAAGA TGCTTGGCTA GGATCTGCGA ACTACACCAA TCTTCTCTA
451 CGTTAGATA ATAATCTCAT TCTAGGAATG CATAGCTCGG AGCTCTGTGA
501 TCTCATATC ACAAAATACCT CTGGAGACTT TTCTATAAAG GATCAAACAG
551 GAAAGTATT TGTTCTCCT CAAGATCGTA AAATTGCAAT ACAAGCTGTA
601 CTCGAAAAAA TCCAGACAGC TCAGAAAACC ATCCAAGTTG CTATGTTGC
651 TCTGACCCAC TCGGAGAGTT TTCAAGCCTT ACATCAAGCA AAACAAACGAG
701 GAATCCATGT AGATATTATC ATTGATAGAA GTCATAGCAA ACTTACTTTT
751 AAGCAATTAC GACAATTAAA TATCAATAAA GACTTTGTTT CTATAAATAC
801 CGCACCCCTGT ACTCTTCACC ATAAGTTTG AGTTATAGAT AATAAAACTC
851 TACTTGCAGG ATCTATAAAAT TGGTCTAAAG GAAGATTCTC CTTAAATGAT
901 GAAAGCTTGA TCATACTGGA AACCTGACC AAACAAACAA ATCAGAAACT
951 TCGAATGATT TGGAAAGATC TAGCTAAGCA TTCAGAACAT CCTACAGTAG
1001 ACAGATGAAGA AAAAGAAATT ATAGAAAAAA GTCTTCCAGT AGAAGAGCAA
1051 GAAGCAGCGT GA

```

The PSORT algorithm predicts periplasmic (0.934).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 75A) and also as a
25 GST-fusion. The GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 75B) and for FACS analysis.

These experiments show that cp6607 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 76

30 The following *C.pneumoniae* protein (PID 4376624) was expressed <SEQ ID 151; cp6624>:

35

```

1 MDAKMGYIFK VMRWIFCFVA CGITFGCTNS GFQNANSRPC ILSMNRMIHD
51 CVERVVGPNRL ATAVLIKGSN DPHAYEMVKD DKDKIAGSAV IFCNGLGLEH
101 TLSLRKHLEN NPNSVKLGER LIARGAFVPL EEDGICDPHI WMDLSIWKEA
151 VIEITEVLIE KFPEWSAEFK ANSEELVCEM SILD SWAKQC LSTIPENLRY
201 LVSGHNAFSY FTRRYLATPE EVASGAWRSR CISPEGLSPE AQISVRDIMA
251 VVDYINEHDV SVVFPEDTLN QDALKKIVSS LKKSHLVRLA QKPLYSNDNV
301 DNYFSTFKHN VCLITEELGG VALECQR*

```

The cp6624 nucleotide sequence <SEQ ID 152> is:

40

```

1 ATGGATGCGA AAATGGGATA TATATTAAA GTGATGCGTT GGATTTCCTG
51 TTTCTGGCA TGTGGTATAA CTTTTGGATG TACCAATTCT GGGTTTCAGA
101 ATGCAAATTC ACCTCCTTGT ATACTATCCA TGAATCCCAT GATTCTATGAT
151 TGTGTTAAA GAGTCGTGGG GAATAGGCTT GCTACCGCTG TTTTGATCAA
201 AGGATCCTTA GACCCCTCATG CGTATGAGAT GGTTAAAGGG GATAAGGACA
251 AGATTGCTGG AAGTGCCTGA ATTCTTGATG ACGGCCCTGGG TCTTGAGCAT
301 ACATTAAGTT TCGGGAAGCA TTAGAAAAT AATCCCATAA GTGTCAGTT
351 AGGGGAGCGG TTGATAGCGC GTGGGGCCCTT TGTTCCTCTA GAAGAAGACG
401 GTATTTGCGA TCCTCATATC TGGATGGATC TTTCTATTG GAAGGAAGCT
451 GTCATAGAAA TTACAGAAGT TCTCATTGAA AAGTCCCTG AATGGTCTGC
501 TGAATTAAA GCAAATAGTG AGGAACCTGT TTGTGAAATG TCTATTTCAG
551 ATTCTTGGGC GAAACAAATGC TTGAGCACAA TCCCTGAAA TTTACGGTAT
601 CTTGTCTCAG GTCATAATGC GTTCAGTTAC TTTACACGTC GCTATTTCAGC
651 TACTCCTGAA GAAGTGGCTT CCGGAGCATG GAGGTCTCGT TGTATTTCCTC
701 CTGAGGGTCT ATCTCCAGAA GCTCAAATCA GTGTTCGTGA TATTATGGCG
751 GTTGTAGATT ATATTAATGA GCATGATGTC AGTGTGGTTT TCCCTGAGGA
801 TACTCTGAAC CAAGATGCGT TGAAAAAAAT TGTTCCTCT CTGAAGAAAA
851 GTCATTTAGT TCGTCTAGCT CAAAACCCT TGTATAGTGA TAATGTGGAC
901 GACAATTATT TTAGCACCTT TAAACATAAT GTCTGCCTTA TCACAGAAGA

```

1 ATGTTCGTAA TGAAAAAACT TGTCCGTCTA TGCCTAGTTC TTCTTTCTTT
 51 ACTTCCGAAT CTATTATTTT CTTCGGATCT TTTACGAGAA GAGGCCATCA
 101 AAAAGATGAT GGACAAGCTG ATCGAGTATC ATGTCGATGC TCAAGAGGTT
 151 TCTACGGATA TACTCTCGCG TTCTTTATCT AGTTACATTC AATCTTTGAA
 201 TCCTCATAAA TCTTATCTTT CAAACCAAGA GGTTGCAGTT TTTCTACAGT
 251 CTCGGGAAAC AAAGAACAGT CTCTTAAAGA ATTATAAGGC AGGCAACTTT
 301 GCTATTTATC GCAACATCAA TCAATTAAATT CATGAGAGTA TTCTTCGTGC
 351 CAGGCAGTGG AGAAACGAAT GGGTTAAAGA TCCAAAAGAG CTTGTATTGG
 401 AGGCATCCTC ATATCAGATA TCGAAGCAAC CTATGCAATG GAGCAAATCT
 451 TTAGACGAAG TGAAGCAGAG ACAACGCGCT CTACTCCTTT CCTATCTTTC
 501 TTACATCTT GCTGGAGCTT CTTCTCTCG TTATGAGGGT AAAGAAGAGC
 551 AGCTTGCTGC TCTGTGTCTA CGTCAAATCG AGAACCATGA GAATGTATAT
 601 TTAGGTATCA ACGATCATGG TGTGCTATG GATCGGGATG AAGAAGCCTA
 651 CCAATTCCAT ATCCGTGTTG TTAAAGCTTT AGCTCATAGC TTAGATGCAC
 701 ATACGGCGTA TTTCAGTAAG GACGAAGCGT TGGCGATGCG AATCCAACTA
 751 GAAAAAGGCA TGTGTGGAAT TGGTGTGTT CTGAAGGAAG ATATTGATGG
 801 AGTTGTTGTT AGAGAAATCA TTCTGGGGG ACCTGCGGCT AAATCTGGGG
 851 ATCTTCAGCT TGGAGATATC ATCTATCGGG TGGATGGCAA CGATATCGAG
 901 CATCTTCTT TCCGCGGTGT TTTAGATTGT TTACGTGGAG GTCATGGCTC
 951 TACTGTAGTC TTAGATATCC ATCGTGGGG AAGCGATCAT ACGATCGCCT
 1001 TGAGAAGGGGA GAAAATCCTT TTAGAAGACC GTCGTGTGGA TGTTTCCSTAT
 1051 GAGCCTTATG GAGATGGTGT GATGGGGAAA GTTACGTTAC ATTCTTTTTA
 1101 TGAAGGAGAA AATCAGGTTT CTAGTGAACA AGATCTACGT CGAGCGATTG
 1151 AGGGATTAAA GGAGAAGAAC CTTCTTGGAT TAGTTTAGA TATCCGAGAA
 1201 AATACGGGTG GATTTTATC TCAAGCGATC AAAGTTCTG GTTTATTAT
 1251 GACCAATGGC GTTGTGTTG TATCTCGCTA TGCTGATGGT ACCATGAAGT
 1301 GCTACCGCAC AGTATCTCCT AAAAAATTCT ATGATGGTCC TTTGGCTATT
 1351 TTAGTATCTA AAAGTCCGC ATCAGCAGCG GAGATTGTAG CACAAACTCT
 1401 CCAAGATTAT GGAGTTGCTT TAGTTGTTGG AGATGAGCAG ACCTATGGGA
 1451 AGGGAACGAT TCAGCATCAA ACAATTACTG GAGATGCCCTC TCAGGACGAT
 1501 TGTGTTAAGG TTACTGTAGG GAAATATTAT TCCCCTCTG GGAAATCGAC
 1551 TCAACTTCAG GGAGTAAAAT CCGATATTTC AATTCCTTCT CTCTATGCTG
 1601 AAGATCGTCT AGGAGAGCGT TTCTTAGAGC ATCCCTTACG TGCAAGATTGC
 1651 TGTGATAATG TACTTCACGA TCCTCTCACG GACTTGGATA CTCAAACACG
 1701 TCCTGGTTT CAAAAAATACT ATCTTCCCTAA TCTACAAAAG CAAGAGACTC
 1751 TTGGAGAGA GATGCTACCT CAGCTTACGA AAAACAGTGA GCAAAGGCTT
 1801 TCTGAGAATT CGAATTTCGA GGCATTTTG TCGCAGATAA AATCATCTGA
 1851 AAAAAACGGAC CTATCCTATG GTTCCAATGA TTACAATTG GAAGAGTCGA
 1901 TAAACATTTT GAAGGACATG ATTTTATTAC AACAGTGTAG AAAATAA

40 The PSORT algorithm predicts periplasmic (0.932).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 78A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 78B) and for FACS analysis.

45 These experiments show that cp6847 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 79

The following *C.pneumoniae* protein (PID 4376969) was expressed <SEQ ID 157; cp6969>:

1 MRLFSLGLTI LFFSLALSSC CGYSILNSPY HLSSLGKSL QERIFIAPIK
 51 EDPHGQLCSA LTYELSKRSF AISGRSSCAG YTLKVELLNG IDKNIGFTYA
 101 PNKLGDKTHR HFIVSNEGRL SLSAKVQLIN NDTQEVLIDQ CVARESVDFD
 151 FEPDLGTTANA HEFALQFEM HSEAIKSARR ILSIRLAETI AQQVYYDLF*

A predicted signal peptide is highlighted.

The cp6969 nucleotide sequence <SEQ ID 158> is:

1 ATGAGATTGT TTTCTTCTAGG CACGATTTAT CTTTTTTTTT CTCTAGCACT
 51 TTCGTCATGC TGTGGTTACT CTATTTAAA CAGCCCCGTAT CACTTATCGT
 101 CTTTAGGTAA GTCTTTATTA CAGGAAAGAA TTTTCATTGC TCCCATAAAA

1301 ACGTCACCTC TACTATCCGA CAAACCTGCAG TATTAGCGCG GGGAGATCTT
 1351 GTACTTCGTG ATGGAGTCAC CGTAACCTTC AAGGATCTGA CTCAAAGTCC
 1401 AGGATCCCGC ATCTTAATGG ATGGGGGGAC TACACTTAGT GCTAAAGAGG
 1451 CAAATCTTTG CCTTAATGGC TTAGCAGTAA ATCTCTCCTC TTTAGATGGA
 1501 ACCAACAAAGG CAGCTTTAAA AACAGAACGCT GCAGATAAAA ATATCAGCCT
 1551 ATCGGGAAACG ATTGCGCTTA TTGACACCGA AGGGTCATTC TATGAGAATC
 1601 ATAACTTAAA AAGTGCTAGT ACCTATCCTC TTCTTGAAC TACCACCGCA
 1651 GGAGCCAACG GAACGATTAC TCTGGGAGCT CTTTCTACCC TGACTCTTCA
 1701 AGAACCTGAA ACCCACTACG GGTATCAAGG AAACCTGGCAG TTGTCTTGGG
 1751 CAAATGCAAC ATCCCTAAAA ATAGGAAGCA TCAACTGGAC CCGTACAGGA
 1801 TACATTCTA GTCCCTGAGAG AAAAAGTAAT CTCCCTCTAA ATAGCTTATG
 1851 GGGAAACTTT ATAGATATAC GCTCGATCAA TCAGCTTATA GAAACCAAGT
 1901 CCAGTGGGA GCCTTTGAG CGTAGCTAT GGCTTTCAGG AATTGCGAAT
 1951 TTCTTCTATA GAGATTCTAT GCCCACCCGC CATGGTTTCC GCCATATCAG
 2001 CGGGGGTTAT GCACTAGGGA TCACAGCAAC AACTCCTGCC GAGGATCAGC
 2051 TTACTTTTGC CTTCTGCCAG CTCTTGCTA GAGATCGCAA TCATATTACA
 2101 GGTAAAGAAC CACGGAGATAC TTACGGTGCC TCCTTGTATT TCCACCATAC
 2151 AGAAGGGCTC TTGACATCG CCAATTCTCT CTGGGGAAAA GCAACCCGAG
 2201 CTCCCTGGGT GCTCTCTGAG ATCTCCCAGA TCATTCTTT ATCGTTCGAT
 2251 GCTAAATTCA GTTATCTCCA TACAGACAAC CACATGAAGA CATATTATAC
 2301 CGATAACTCT ATCATCAAGG GTTCTTGGAG AAACGATGCC TTCTGTGCG
 2351 ATCTTGGAGC TAGCCTGCCT TTTGTTATTT CCGTTCCGTA TCTTCTGAAA
 2401 GAAGTCGAAC CTTTGTCAA AGTACAGTAT ATCTATGCCG ATCAGCAAGA
 2451 CTTCTACGAG CGTCATGCTG AAGGACGCGC TTTCATAAAA AGCGAGCTTA
 2501 TCAACGTAGA GATTCTATA GGCGTCACCT TCGAAAGAGA CTCAAAATCA
 2551 GAAAAGGGAA CTTACGATCT TACTCTTATG TATATACTCG ATGCTTACCG
 2601 ACGCAATCCT AAATGTCAA CTTCCCTAA AGCTAGCGAT GCTAACTGGA
 2651 TGGCCTATGG TACCAACCTC GCACGACAAG GTTTTCTGT TCGTGCTGCG
 2701 AACCATTTCC AAGTGAACCC CCACATGGAA ATCTTCGGTC AATTGCTTT
 2751 TGAAGTACGA AGTTCTTCAC GAAATTATAA TACAAACCTA GGCTCTAAGT
 2801 TTGTTTCTA G

The PSORT algorithm predicts inner membrane (0.187).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 77A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 35 77B) and for FACS analysis.

The cp6728 protein was also identified in the 2D-PAGE experiment.

These experiments show that cp6728 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 78

40 The following *C.pneumoniae* protein (PID 4376847) was expressed <SEQ ID 155; cp6847>:

1 MFVMKKLVRL CVVLLSLLPN VLFFSDLLRE EGIKKMDKL IEYHVDAQEV
 51 STDILSRSLN SYIQSFDPHK SYLSNQEVAV FLQSPETKKR LLKNYKAGNF
 101 AIYRNINQLI HESILRARQW RNEWVKNPKE LVLEASSYQI SKQPMQWSKS
 151 LDEVVKQRQA LLLSYLSLHL AGASSSSRYEG KEEQLAALCL RQIENHENVY
 201 LGINDHGVM DRDEEAYQFH IRVVKALAHs LDAHTAYFSK DEALAMRIQL
 251 EKGMCIGVV LKEDIDGVVV REIIPGGPAA KSGDLQLGDI IYRVDGKDIE
 301 HLSFRGVLDL RLRGGHGSTVV LDIHRGESDH TIALRREKIL LEDRRVDVSY
 351 EPYGDGVIGK VTLHSFYEGE NQVSSEQDLR RAIQGLKEKN LLGLVLDIRE
 401 NTGGFLSQAI KVSGLFMTNG VVVVSRYADG TMKCYRTVSP KKFYDGPLAI
 451 LVSKSSASAA EIVAQTLQDY GVALVVGDEQ TYKGTIQHQ TITGDASQDD
 501 CFKVTVGKYY SPSGKSTQLO GVKS DILIPS LYAE DRLGER FLEHPLPADC
 551 CDNVLHDPLT DLDTQTRPWF QKYLYPNLQK QETLWREMLP QLTKNSEQR
 601 SENSNFQAF SQIKSSEKTD LSYGSNDLQL EESINILKDM ILLQOCR*

A predicted signal peptide is highlighted.

55 The cp6847 nucleotide sequence <SEQ ID 156> is:

-120-

These experiments show that cp7109 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 81

The following *C.pneumoniae* protein (PID 4377110) was expressed <SEQ ID 161; cp7110>:

```

5      1 MAAIKQILRS MLSQSSLWMV LFSLYSLSGY CYVITDKPED DFHSSSAVKW
      51 DHWGKTTLRS LSNKKASAKA VSGIGATTVG FIKDTWSRTY AVRWNWYWGTK
     101 ELPPTSSWVKK SKATGISSDG SIIAGIVENE LSQSFAVTWK NNEMYLLPST
     151 WAVQSKAYGI SSDGSVIVGS AKDAWSRTFA VKWTGHEAQV LPVGWAVKSV
    10  201 ANSVSANGSI IVGSVQDASG ILYAVKWEGRN TITHLGTLGG YSAIAKAVSN
    251 NGKVIVGRSE TYYGEVHAF C HKNGVMSDLG TLGGSYSAAK GVSATGKVIV
     301 GMSTTANGKL HAFKYVGGRM IDLGEYSWKE ACANAVSIDG EIIVGVQSE*

```

A predicted signal peptide is highlighted.

The cp7110 nucleotide sequence <SEQ ID 162> is:

```

15     1 ATGGCAGCTA TAAAACAAAT TTTACGTTCT ATGCTATCTC AGAGTAGCTT
      51 ATGGATGGTC CTATTTCTAT TATATTCTCT ATCTGGTTAT TGCTATGTAA
     101 TTACAGACAA ACCAGAACAT GACTTCCATT CTTCATCCGC AGTAAAATGG
     151 GATCATTGGGG GAAAGACAAAC TCTCTCAAGA TTATCAAATA AAAAAGCCTC
     201 TGCAAAAGCT GTTTCAGGAA CTGGTGCTAC AACTGTCGGC TTTATAAAAG
     251 ACACCTGGTC TCGAACATAC GCAGTAAGAT GGAATTATG GGGGACCAAA
    20  301 GAACTCCCTA CCAGCTCATG GGTTAAAAAA TCAAAAGCAA CAGGAATCTC
     351 CTCTGATGGGG TCTATAATCG CGGGGATTGT CGAGAAATGAG CTTTCTCAAA
     401 GTTTCGCGAGT CACATGGAAA ACAATGAAA TGTATTGCT CCCTTCCACA
     451 TGGGCAGTGC AATCTAAAGC GTATGGAATT TCTTCTGATG GCTCTGTTAT
    25  501 TGTAGGGAGT GCTAAGGATG CTTGGTCGCG AACTTTCGCT GTGAAGTGG
     551 CGGGACACGA GGCTCAGGTG TTACCAGTAG GCTGGGCTGT CAAATCTGTA
     601 GCGAATTCTG TATCTGCCAA TGGATCTATA ATTGTAGGGT CTGTACAAGA
     651 CGCCTCTGGG ATTCTTTATG CTGTAAAGTG GGAAGGGAAC ACTATTACAC
     701 ATCTAGGAAC TTTAGGAGGC TATTCTGCCA TTGCAAAAGC TGTATCCAAT
     751 AATGGCAAGG TCATTGTAAG GAGATCCGAA ACATATTATG GAGAGGTCCA
    30  801 TGCTTTCTGT CATAAGAATG CGCTCATGTC AGACCTCGGC ACCCTCGGAG
     851 GATCTTATTG TGCAGCTAAG GGAGTCTCTG CAACTGGAAA AGTTATTGTC
     901 GGTATGTCCA CAACAGCAAA TGGGAAATTG CATGCCCTTA AATATGTCGG
     951 TGGAAAGAATG ATCGACTTAG GAGAGTATAG CTGGAAAGAA GCCTGTGCAA
    1001 ACGCTGTTTC TATTGATGGA GAAATTATTG TTGGAGTCGA ATCAGAAATAA

```

35 The PSORT algorithm predicts outer membrane (0.827).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 81A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 81B) and for FACS analysis.

40 These experiments show that cp7110 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Figure 191 shows a schematic representation of the structural relationships between of cp7105, cp7106, cp7107, cp7108, cp7109 and cp7110, each of which is identified herein. These six proteins may be grouped in a new family of related outer membrane-associated proteins. These proteins have a repeat structure in common (cf. the pmp family).

Example 82

The following *C.pneumoniae* protein (PID 4377127) was expressed <SEQ ID 163; cp7127>:

```

1 MVFFRNSLLH LVALSGMLCC SSGVALTIAE KMASLEHSGR GADDYEGMAS

```

151 GAAGATCCTC ATGGTCAGCT CTGCTCAGCT CTAACATTATG AGCTTAGTAA
 201 GCGTTCTTT GCTATCTCTG GAAGGAGTTC TTGCGCAGGC TATACTCTTA
 251 AAGTAGAGCT TCTGAATGGT ATTGACAAGA ATATAGGTTT TACGTATGCC
 301 CCAAATAAAC TCGGAGATAA GACTCACAGG CATTATAG TCTCTAATGA
 351 AGGCAGACTA TCACTATCTG CAAAAGTACA GCTTATCAAT AATGACACTC
 401 AAGAAGTCCT TATAGACCAA TGTGTTGCTC GAGAGTCTGT AGACTTTGAC
 451 TTTGAGCCCTG ACTTAAAGAAC AGCAAAACGCT CATGAATTG CTTTAGGCCA
 501 ATTTGAAATG CATAGTGAAG CCATAAAAAG TGCTCGCCGT ATACTATCTA
 551 TACGCCCTAGC CGAGACGATT GCTCAACAGG TATACTATGA CCTTTTTGTA

- 10 The PSORT algorithm predicts inner membrane (0.126).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 79A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 79B) and for FACS analysis.

- 15 These experiments show that cp6969 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 80

The following *C.pneumoniae* protein (PID 4377109) was expressed <SEQ ID 159; cp7109>:

1 MKKTCCQNYR SIGVVFSVVL FVLTTQTLFA GHFIDIGTSG LYSWARGVSG
 51 DGRVVVGYEG GNAFKYVDGE KFLLEGLVPR SEALVFKASY DGSVIIGISD
 101 QDPSCRAVKW VNGALVDLGI FSEGMSFAE GVSSDGKTIV GCLYSDDTET
 151 NFAVKWDETG MVVLPNLPED RHSCAWDASE DGSVIVGDAM GSEEIAKAVY
 201 WKDGEQHLLS NIPGAKRSSA HAVSKDGSFI VGEFISEENE VHAFVYHNGV
 251 IKDIGTLGGD YSVATGVSRD GKVIVGHSTR TDGEYRAFKY VDGRMIDLGT
 301 LGGSASFAFG VSDDGKTIVG KFETELGECH AFYLDD*

- 25 A predicted signal peptide is highlighted.

The cp7109 nucleotide sequence <SEQ ID 160> is:

1 ATGAAAAAGA CATGTTGCCA AAATTACAGA TCGATAGGGC TTGTGTTCTC
 51 TGTGGTACTT TTCGTTCTTA CAACACAGAC GCTGTTGCA GGACATTTTA
 101 TTGATATTGG AACTTCTGGA TTATATTCTT GGGCTCGAGG TGTATCTGGA
 151 GATGGCCGCG TTGTCGTAGG TTATGAAGGT GGCAATGCAT TTAAATATGT
 201 TGATGGTGAG AAATTTCTGT TAGAAGGTTT GGTCCCGAGA TCCGAGGCCT
 251 TGGTATTAA AGCTTCTTAT GATGGCTCTG TAATTATAGG AATCTCGGAT
 301 CAAGATCCGT CTTGCCGCG TGTGAAGTGG GTAAACGGTG CACTTGTGA
 351 TCTTGAATA TTTTCTGAGG GAATGCAATC TTTTGCAGAG GGTGTTCCA
 401 GTGATGGAAA GACGATTGTA GGGTGCCTAT ATAGTGATGA TACAGAGACA
 451 AACTTTGCTG TGAAGTGGGA TGAAACAGGA ATGGTTGTC TCCCTAACTT
 501 ACCAGAAGAT CGACATTCTT GCGCTTGGGA TGCCTCTGAA GATGGCTCTG
 551 TGATTGTAGG GGACGCCATG GGTAGCGAGG AAATTGCCAA GGCAGTGTAC
 601 TGGAAGGACG GTGAACAACA TCTGTTTCT AATATCCCAG GAGCTAAAG
 651 ATCGTCAGCA CATGCAGTT CTAAAGATGG ATCTTTTATC GTAGGGAGT
 701 TCATCAGTGA AGAAAATGAA GTTCATGCCT TTGTTTATCA CAACGGTGT
 751 ATCAAAGATA TCGGGACTTT AGGAGGAGAT TACTCTGTAG CAACTGGAGT
 801 TTCTAGGGAT GGTAAGGTCA TCGTGGTCA TTCTACAAGA ACAGATGGTG
 851 AATACCGTGC ATTTAAATAT GTGGATGGAA GAATGATAGA TTTGGGACT
 901 TTAGGAGGTT CAGCATCTT TGCTTTGGT GTTTCTGACG ATGGCAAAAC
 951 AACCGTAGGA AAATTTGAAA CAGAGCTAGG AGAATGTCA GCCTTATCT
 1001 ACCTTGATGA TTAG

The PSORT algorithm predicts outer membrane (0.887).

- 50 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 80A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 80B) and for FACS analysis.

-122-

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 82A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 82B) and for FACS analysis.

These experiments show that cp7127 is a surface-exposed and immunoaccessible protein, and that it
5 is a useful immunogen. These properties are not evident from the sequence alone.

Example 83

The following *C.pneumoniae* protein (PID 4377133) was expressed <SEQ ID 165; cp7133>:

```

1  MOPFIFTLLE LTSLVSLVAF DAANARKRCA CAQTIERGEN FFSIKRSACA
51 EIEYQEKSRRH ASAIERISKD KGKVTPKQIA KVATKKQRY RLLQVPPFSRP
101 PNSRNYLVA LLSEPPECYS DTASWYAIFI RLLRRAYVDT GNVPPGSEYAA
151 IANALISNKQ EILERGAQLG PDVIETLTLP EEQAEIFYKM LKGSSNSQSL
201 LNFLHYEEKS LGHCKLNLIF MDPLLLEAVL DHPDAYRETS LLRDGIWEAV
251 KRQEHAIQEH GQAAALELFK TRTDFRLELR DKMQLLLSRY DLLPLLNKKM
301 FDYTLGSAGD YLFLVDPDTK AISRCRCPSK SIKL

```

15 A predicted signal peptide is highlighted.

The cp7133 nucleotide sequence <SEQ ID 166> is:

```

1  ATGCAACCTT TTATCTTAC TTTACTGTGC TTGACATCCTT TGGTTCTTT
51 AGTCGCCCTT GATGCTGCGA ATGCTCGTAA ACGTTGTGCC TGTGCTCAAA
101 CTATAGAACG TGGAGAGAAC TTCTTTTCCA TAAAACGCTC TGCTTGTGCT
151 GAAATCGAAT ATCAAGAAAAA ATCTCGCCAC GCCTCAGCAA TTGAAAGAAT
201 CTCAAAAGAT AAAGGCAAG TCACTCCAAA GCAGATTGCG AAAGTAGCTA
251 CTAAGAAAAA GCAAAGATAC CGTTTATTGC AGGTTCCCTT TTCAAGGCCT
301 CCGAATAACT CAAGGTATAA CCTCTATGCT TTGCTTAGTG AACCTCCCGA
351 ATGCTATAGC GATACAGCAT CATGGTATGC TATTTTATT CGGTTACTTC
401 GACGTGCTTA TGTAGACACG GGAAATGTAC CTCCCTGGATC TGAGTATGCC
451 ATCGCTAATG CTTTGATAAG TAACAAACAA GAGATTTAG AGAGGGGAGC
501 GCAGCTTGGG CCCGATGTTA TTGAAACTCT AACATTGCCT GAGGAACAAG
551 CCGAGATTTT TTATAAAATG CTCAAAGGGT CGTCAAACTC TCAGTCGCTA
601 CTGAATTTC TGCATTATGA AGAGAAAAGC TTAGGCCACT GTAAAGCTAAA
651 TCTGATCTTC ATGGATCCCC TACTGTTAGA AGCTGTTCTA GATCATCCCG
701 ATGCTTATAG GGAAACGTCG CTCCCTGCGCG ATGGCATTG GGAAGCGGTG
751 AAGCGTCAAG AACATGCCAT CCAAGAACAT GGCCAGGCAG CTGCTTTGGA
801 GCTTTTTAAA ACACGCCACCG ACTTCCGCCT GGAGCTGCGA GATAAGATGC
851 AGTTACTTCT AAGTCGATAC GATTTGCTCC CCTTATTAAA TAAAAAAATG
901 TTCGACTACA CCTTAGGAAG TGCCGGAGAT TACTTATTG TGTTAGACCC
951 AGATACTAAG GCAATTCTC GATGTCGCTG CCCTCAAAG AGTATTAAAT
1001 TATAA

```

The PSORT algorithm predicts outer membrane (0.92).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 83A) and also in
40 his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 83B) and for FACS analysis.

These experiments show that cp7133 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 84

45 The following *C.pneumoniae* protein (PID 4377222) was expressed <SEQ ID 167; cp7222>:

```

1  MNRRDMVITA VVVNAILLVA LFVTTSKRIGV KDYDEGFRNF ASSKVTQAVV
51 SEEKVIEKPV VAEVPSRPIA KETLAAQFIE SKPVIVTTPP VPVVSETPEV

```

-121-

5 51 FNANMREYSL QLSKLYEEAR KLRASGTEDE ALWKDLIRRI GEVRGYLREI
 101 EELWAAEIRE KGGNLEDYAL WNHPETTIYN LVTDYGTEDS IYLIHQEIGA
 151 IKIATLSKFV VPKESFEDCL TQILSRLGIG VRQVNSWIKE LYMMRKEGCS
 201 VAGVFSSRKD LEALPETAYI GFVLNSNVDA HTNQHVLKKF INPETTHVDV
 251 IAGRIVWIFGS AGEVGELLKI YNFVQSESIR QEYRVIPLTQ IDPGEMISIL
 301 NAAFREDLTK DVSEESLGLR VVPLQYQGRS LFLSGTAALV QQALTIREL
 351 EEEGIENPTDK TVFWYVNKHQS DPQELAALLS QVHDVFSGEN KASVGAADGC
 401 GSQLNASIQQI DTTVSSSAKD GSVKYGNFIA DSKTGTLIMV VEKEVLPRIQ
 451 MLLKKLDVPK KMVRIEVLF ERKLAHEQKS GLNLLRLGEE VCKKGCSPSV
 501 SWAGGTGILE FLFKGSTGSS IVPGYDLAYQ FLMAQEDVRI NASPSVVTMN
 551 QTPARIAAVD EMSIAVSSDK DKAQYNRAQY GIMIKMLPVI NVGEEDGKSY
 601 ITLETDTIFD TTGKNHDDR P DVTRRNITNK VRIADGETVI IGGLRCKQMS
 651 DSHDGIPFLG DIPGIGKLFG MSSTSDSLTE MFVFTPKIL ENPVHQERK
 701 EEALLSSRPG EREEYYQALA ASEAAARAHH KKLEMFPASG VSLSQVERQE
 751 YDGC*

A predicted signal peptide is highlighted.

The cp7127 nucleotide sequence <SEQ ID 164> is:

20 1 ATGGTTTTTT TCCGTAATT C TTTACTGCAT TTAGTTGCC TATCCGAAT
 51 GCTCTGTTGT TCTTCTGGAG TGGCTTTAAC GATAGCCGAG AAGATGGCTT
 101 CTTTAGAGCA CTCGGGGAGA GGAGCAGACG ATTATGAGGG GATGGCTTCG
 151 TTTAATGCCA ATATGAGGG GTATAGCCTT CAGCTGAGCA AGTTGTATGA
 201 GGAAGCACGA AAGCTACCGC CTTCTGGAAC TGAGGATGAA GCTCTGTGGA
 251 AGGACTTAAT TCGACGGATT GGTGAGGTGC GAGGCTATCT TCGAGAGATC
 301 GAGGAGCTTT GGGCTGCAGA AATTCTGTGAG AAAGGGGGCA ATCTCGAGGA
 351 CTACGCCCTC TGGAATCACC CAGAGACTAC GATTTACAAT CTTGTTACCG
 401 ATTACGGAAC CGAAGACTCT ATTATTGTA TTCCCTCAAGA AATCGGAGCG
 451 ATTAAAATCG CAACCTTATC GAAATTGTA GTTCCCTAAAG AGTCTTTCGA
 501 AGACTGTCTC ACTCAGATCC TATCTCGCTT AGGTATTGGC GTGCGTCAGG
 551 TCAATTCTTG GATTAAGGAA CTTTATATGA TGCGTAAGGA GGGCTGCAGT
 601 GTTGCTGGAG TTTTTCTTC CAGAAAAGAT TTAGAGGCAG TCCCCAGAAC
 651 AGCCTATATT GTTTTTGTAT TGAATTGCAA CGTAGATGCG CATACCAATC
 701 AACATGTCTT AAAAAAGTC ATTAAACCTG AAACAACGCA TGTAGATGTG
 751 ATTGCAAGGAC GTGTGTGCGAT TTTTGGTTCT GCGGGGGAAAG TCGGGAGCT
 801 TCTGAAGATT TATAATTTCG TGCAGTCGGA GAGCATACTG CAAGAGTATC
 851 GGGTGAATTCC CTTAACTAAG ATCGATCCAG GGGAGATGAT TTCCATTCTC
 901 AACGCAGCAT TTCGTGAGGA TCTGACTAAA GATGTTAGTG AAGAATCTTT
 951 AGGCCCTCGT GTAGTTCCCT TACAGTATCA AGGGGGTTTCG TTGTTTTAA
 1001 GTGGAACCGC GGCCTTGTG CAGCAAGCGC TGACTCTCAT TCGAGAGCTT
 1051 GAAGAAGGGG TTGAGAACCC TACGGATAAA ACAGTATTG GGTATAACGT
 1101 CAAGCACTCC GATCCCCAAG AGTTGGCGGC ATTGCTTCTC CAAGTCATG
 1151 ATGCTTCTC TGGCAGAAT AAGGCGAGTG TCGGAGCTGC AGATGGATGT
 1201 GGGTCGCAAT TAAATGCCCTC GATCCAAATT GATACTACAG TAAGTCTTC
 1251 TCGCAAAGAT GGCTCAGTGA AGTACGGAAA CTTCATCGCG GATTCTAAGA
 1301 CAGGAACCTCT GATTATGGTG GTTGAGAAAG AAGTTCTTC ACgtATTCAg
 1351 ATGCTACTTA AGAAACTAGA TGTCCTAAA AAGATGGTCC GTATCGAGGT
 1401 GCTGTTATTG GAAAGAAAAT TGGCACATGA GCAGAAATCT GGGTTAAATC
 1451 TTCTACGTCT TGGTGAGGAA CTTTGTAAAAA AAGGGTGCAG TCCTCTGTG
 1501 TCTTGGGCCG GGGGTACTGG CATACTAGAA TTTTTATTTA AAGGAAGTAC
 1551 GGGATCTTCG ATAGTTCCGT GTTATGATCT CGCCTATCAA TTTTTATGG
 1601 CTCAGAGGA CGTTCCGATT AATGCGAGTC CTTCTGTAGT TACTATGAAC
 1651 CAAACCCCAG CACGGATTGC TGTGTTGAT GAAATGTCAA TAGCCGTGTC
 1701 TTCAGATAAA GATAAAGCGC AATACAATCG TGCGCAGTAC GGTATCATGA
 1751 TAAAAATGCT CCCCGTAATT AATGTGGGAG AGGAAGACGG AAAAGTTAC
 1801 ATTACTTTAG AGACAGACAT CACCTTGTGAT ACTACGGGAA AAAATCATGA
 1851 TGATCGTCT GATGTTACAA GGGCTAATAT TACTAATAAG GTGCCATTG
 1901 CTGACGGAGA GACTGTGATT ATTGGAGGTT TGCGTTGCAA ACAGATGTCA
 1951 GATTCTCATG ATGGCATTCC TTTCCCTGGA GACATTCTG GTATAGGGAA
 2001 GTTATTGGA ATGAGTTCCA CATCAGACAG TCTCACGGAG ATGTTGTAT
 2051 TTATCACTCC GAAGATCCTA GAAAATCCTG TAGAGCAACA AGAACGTAAA
 2101 GAAGAAGCTT TACTCTCTC GCGCCCTGGA GAGAGAGAAG AATACTATCA
 2151 GGCTTCTGAGCA GCTAGTGTAGG CTGCGACAG AGCAGCTCAT AAAAATTAG
 2201 AGATGTTCCC GGCATCAGGA GTATCTTTAT CTCAGGTAGA GAGGCAAGAA
 2251 TACGATGGCT GCTAG

The PSORT algorithm predicts periplasmic (0.920).

-124-

401 CCTATGCTAT TGGAGGACTC GCTGCAAAC GCCTGAATGG GTATTCTGGA
 451 TCATCGAAAA TCTTCGTTGC CGAAGCCGAT GAAAGTGTG GGTCTTTAAA
 501 GCACTACACT CCCCCTGCAG TAGTCATTAC AAATATAGAT AATGAACATT
 551 TGAATAATTA CGCTGGAAAT CTTGATAACC TGTTTCAGGT AATCCAGGAC
 601 TTCTCTAGAA AAGAACAGA TCTCAATAAG GTATTCTATA ACGGGGATTG
 651 TCTTATTTTG AAAGGAAATG TCCAAGGGAT TTCTTATGGA TATTACCCAG
 701 AATGTCAATT GCATATCGTT TCCTATAATC AAAAGGCATG GCAATCTCAC
 751 TTTCTCTTTA CCCTTTTAGG CCAGGAGTAT CAAGACATG AGCTCAATCT
 801 CCCTGGACAA CATAACGCTG CAAATGCAGC AGCAGCCTGT GGAGTTGCTC
 851 TTACCTTGTG CATAGACATA AACATCATTC GAAAAGCTCT CAAAAAAATTG
 901 TCGGGAGTTC ATCGACGTCT AGAAAGAAAA AATATATCCG AAAGCTTCT
 951 TTTCTTAGAA GATTATGCTC ATCATCCTGT AGAGGTTGCA CATAACCTGC
 1001 GCTCTGTGCG TGATGCTGTG GGTTTGCAGA GAGTCATCGC AATTTTCAG
 1051 CCACATCGAT TCTCTCGTT AGAAGAGTGC TTACAAACCT TCCCCAAAGC
 1101 TTCTCAAGAA GCTGATGAAG TCATACTTAC AGATGTCTAT AGTGCCTGGAG
 1151 AAAGTCTTAG AGAGTCTATC ATTCTTCCG ACCTTGCAGA ACAGATTCTG
 1201 AAGTCTTCTT ATGTCCATTG TTGTTATGTT CCCATGGAG ACATCGTAGA
 1251 TTATCTACGA AACTACATTG GCATTCATGA TGTCGTGTT TCTCTAGGAG
 1301 CTGGAAATAT CTATACTATT GGAGAGGCTT TAAAAGACTT TAACCTAAA
 1351 AAATTATCCA TAGGACTCGT CTGTGGAGGG AAATCTTGC G AACACGATAT
 1401 TTCTCTACTT TCTGCTCAAC ATGTCCTAA ATATATTCT CCTGAATTCT
 1451 ATGATGTGAG TTACTTCATC ATAATCGTC AGGGCTTATG GAGAACAGGA
 1501 AAGGATTTTC CTCATCTTAT TGAGAGACT CAAGGGGATT CGCCACTTT
 1551 TTCTGAAATC GCTTCAGCTT TAGCAAAAGT CGACTGTTG TTTCCCGTGC
 1601 TCCATGGCCC ATTTGGAGAG GATGGTACGA TCCAGGGATT TTTTGAATC
 1651 TTAGGAAAAC CTTATGCCGG ACCCTCACTA TCTTTAGCAG CAACTGCAAT
 1701 GGATAAGCTG TTAACAAAAC GAATTGCATC AGCAGTGGGT GTPCCGTGAG
 1751 TCCCTTACCA ACCTTAAAT CTCTGTTCT GGAAACGCAA TCCACAACTA
 1801 TGTATTTCAGA ATCTTATAGA GACATTTCT TTCCCTATGA TTGTAAAAC
 1851 TGCACATTTG GGATCTAGTA TTGGGATATT TTAGTCCGT GATAAAGAGG
 1901 AATTACAAGA AAAGATCTCA GAAGCATTTC TATATGACAC GGATGTGTTT
 1951 GTGGAGGAAA GTCGCTTAGG GTCTCGTGA ATCGAAGTGT CCTGTATCGG
 2001 CCATTCTTCT AGCTGGTATT GTATGGCAGG GCCTAATGAA CGCTGTGGTG
 2051 CTAGTGGTT TATTGATTAT CAAGAGAAAT ATGGATTG TGCGATAGAT
 2101 TGCGBAAAGA TCTCTTTGTA TTTACAGCTC TCACAAGAAT CTTTAGATTG
 2151 TGTAGAGAA CTTGCAGAGC GTGTCTACCG AGCAATGCAA GGAAAAGGTT
 2201 CAGCTCGAAT AGATTTTTC TTGGATGAAG AGGGGAATTA TTGGTTGTCA
 2251 GAGGTCAATC CTATTCCAGG AATGACAGCA GCTAGCCCAT TTTTACAAGC
 2301 TTTTGTTCAC GCAGGATGGA CGCAAGAAC AATTGTAGAT CACTTTATTA
 2351 TAGATGCTCT ACATAAGTTT GATAAGCAGC AGACTATCGA ACAGGCATTC
 2401 ACTAAAGAAC AAGATTAGT TAAAAGATAA

The PSORT algorithm predicts inner membrane (0.16).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 85A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 85B) and for FACS analysis.

These experiments show that cp7225 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 86

The following *C.pneumoniae* protein (PID 4377248) was expressed <SEQ ID 171; cp7248>:

50 1 MKFWLOGCAF VGCLLLTLPC CAARRRASGE NLQQTRPIAA ANLQWESYAE
 51 51 ALEHSKQDHK PICLFFTGSW WCMWCIMQD QILQSSEFKH FAGVHLHMVE
 101 101 VDFPQKNHQEP EEQRQKNQEL KAQYKVTGFP ELVFIDAEKG QLARMGFEPG
 151 151 GGAAYVSKVK SALKLR*

A predicted signal peptide is highlighted.

55 The cp7248 nucleotide sequence <SEQ ID 172> is:

1 ATGAAATTTT GGTTGCAAGG ATGTGCTTTT GTCGGGTTGTC TGCTATTGAC

-123-

101 PTVAVPPQPV RETVKEEQAP YATVVVKGD FLERIARANH TTVAKLMQIN
 151 DLTTTQLKIG QVIKVPTSQD VSNEKTPQTO TANPENYYIV QEGDSPWTIA
 201 LRNHIRLDDL LKMNDLDEYK ARRLKPGDQL RIR*

A predicted signal peptide is highlighted.

- 5 The cp7222 nucleotide sequence <SEQ ID 168> is:

1 ATGAATCGTA GAGACATGGT AATAACAGCT GTCGTAGTGA ATGCTATATT
 51 GCTTGTGGCT CTTTCGCTCA CATCAAAGCG TATTGGCGTC AAGGACTATG
 101 ACGAGGGATT CCGTAATTTC GCTCTAGCA AGGTTACACA AGCAGTAGTT
 151 TCAGAAGAAA AAGTCATAGA AAAGCCTGTA GTCGCAGAAG TGCCCTAGCCG
 201 TCCTATCGCT AAAGAGACTC TAGCTGCACA GTTATTGAA AGTAAGCCGG
 251 TTATTGTAAAC CACACCACCC GTGCCCTGTTG TTAGCGAAC CCCAGAAGTG
 301 CCTACTGTGG CAGTTCCGCC TCAGCCTGTT CGTGAGACAG TAAAAGAGGA
 351 ACAAGCTCCT TATGCTACTG TTGAGTGA AAAAGGAGAT TTTCTCGAAC
 401 GCATTGCGAG AGCAAATCAT ACTACCGTTG CAAAATTGAT GCAGATCAAT
 451 GATCTTACCA CCACCCAAC TAAAATTGGT CAGGTCATCA AAGTCCCTAC
 501 GTCTCAAGAT GTCAGCAACG AAAAAACTCC TCAAACACAG ACCGCAAACC
 551 CTGAAAATTA TTATATCGTC CAAGAAGGGG ATAGCCCTG GACAATAGCA
 601 TTGCGTAACC ATATTGATT GGATGATTG CTAAAATGA ATGATCTCGA
 651 TGAATATAAA GCCCGCGCC TTAAGCCTGG AGATCAGTTG CGCATACGTT
 701 GA

The PSORT algorithm predicts periplasmic (0.935).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 84A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 84B) and for FACS analysis.

- 25 These experiments show that cp7222 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 85

The following *C.pneumoniae* protein (PID 4377225) was expressed <SEQ ID 169; cp7225>:

30 1 MKGTPQYHFI GIGGIGMSAL AHILLDRGYE VSGSDLYESY TIESLKAKGA
 51 RCFSGHDSSH VPHDAVVVYS SSIAPDNVEY LTAIQRSSRL LHRAELLSQL
 101 MEGYESILVS GSHGKTGTSS LIRAIHQEAQ KDPDSAIGGL AANCLNGYSG
 151 SSKIFVAEAD ESDGSLKHYT PRAVVITNID NEHLNNYAGN LDNLVQVIQD
 201 FSRKVTDLNK VFYNGDCPIL KGNVQGISYVG YSPECQLHIV SYNQKAWQSH
 251 FSFTFLGQEY QDIELNLPGQ HNAANAAAAC GVALTFGIDI NIIRKALKKF
 301 SGVHRRLERK NISESFLFLE DYAHHPVEVA HTLRSVRDAV GLRRVIAIFQ
 351 PHFRFSRLEEC LQTFPKAFQE ADEVILTDVY SAGESPRESI ILSDLAEQIR
 401 KSSYVHCCYV PHGDIVDYLR NYIRIHDCVCV SLGAGNIYTI GEALKDFNPK
 451 KLSIGLVCVG KSCEHDISLL SAQHVKSYIS PEFYDVSYFI INRQGLWRTG
 501 KDFPHLIEET QGDSPSSEI ASALAKVDCL FPVLHGPFGE DGTIQGFFEI
 551 LGKPYAGPSL SLAATAMDNL LTKRIASAVG VPVVPYQPLN LCFWKRNPEL
 601 CIQNLIETFS FPMIVKTAHL GSSIGIFLVR DKEELQEKS EAFLYDTDVF
 651 VEESRLGSRE IEVSCIGHSS SWYCMAGPNE RCGASGFIDY QEKYGFDGID
 701 CAKISFDLQL SQESLDCVRE LAERVYRAMQ GKGSARIDFF LDEEGNYWLS
 751 EVNPPIPMTA ASPFLQAFVH AGWTQEQQIVD HFIIDLHKF DKQQTIEQAF
 801 TKEQDLVKR*

The cp7225 nucleotide sequence <SEQ ID 170> is:

50 1 ATGAAGGGAA CTCCTCAGTA TCATTTTATC GGTATCGGTG GTATAGGAAT
 51 GAGCGCTTTA GCTCATATTG TGCTTGATCG TGGCTATGAG GTCTCTGGAA
 101 GCGACTTATA TGAAAGCTAT ACGATCGAAA GCCTGAAAGC TAAAGGTGCG
 151 AGGTGTTCT CAGGCCATGA TTCCTCCCAT GTTCCTCATG ATGCCGTGCG
 201 TGTTTATAGC TCAAGTATAG CCCCTGATAA TGAGAGTAT CTTACCGCTA
 251 TTCAAAGATC ATCACGTCTT CTTCATAGAG CAGAGCTTT GAGTCAGCTT
 301 ATGGAGGGTT ATGAAAGCAT TCTGGTTCA GGAAGCCATG GGAAGACAGG
 351 GACCTCATCT CTAATTCGAG CGATTTTCCA GGAAGCTAG AAAGATCCCT

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 87A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 87B) and for FACS analysis.

These experiments show that cp7249 is a surface-exposed and immunoaccessible protein, and that it
5 is a useful immunogen. These properties are not evident from the sequence alone.

Example 88

The following *C.pneumoniae* protein (PID 4377261) was expressed <SEQ ID 175; cp7261>:

10	MLPISILLFY VILGCL SAYI ADKKKRN VIG WFFAGAFFGF IGLVVLLLP
	SRRNALEKPQ NDPFDNSDLF DDLKKSLAGN DEIPSSGDLQ EIVIDTEKWF
	YLNKDRENVG PISFEELVVL LKGKTYPEEI WWKKGMKD W QRVKDVPSLQ
	151 QALKEASK*

The cp7261 nucleotide sequence <SEQ ID 176> is:

15	1 ATGCTCCCTA TTTCGATT TT ATTATTTTAT GTGATTCTAG GTTGTCTATC
	51 TGCCTACATA GCAGATAAAGA AAAAACGAAA TGTTATTGGC TGGTTTTTG
	101 CAGGAGCATT TTTTGGATT ATTGGTCTAG TTGTCCTTCT TCTTCCTCCT
	151 TCTCGTCGAA ACGCTTTAGA AAAGCCACAA AACGATCCTT TTGATAACTC
	201 CGATCTTTT GATGATTG AAAAAAGTT AGCAGGTAAT GACGAGATAC
	251 CCTCATCGGG ACATCTCAA GAAATCGTTA TCGATACAGA GAAGTGGTTT
20	301 TATTTAAATA AAGATAGAGA AAACGTAGGT CCGATATCTT TTGAGGAGTT
	351 GGTCTGACTT TTAAAGGGAA AAACGTATCC AGAAGAAATT TGGGTATGGA
	401 AAAAGGGAAT GAAAGATG G CAACGAGTGA AGGATGTTCC ATCACTACAA
	451 CAGGTTTGA AAGAACATC AAAATAA

The PSORT algorithm predicts inner membrane (0.848).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 88A). The
25 recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure
88B) and for FACS analysis.

These experiments show that cp7261 is a surface-exposed and immunoaccessible protein, and that it
is a useful immunogen. These properties are not evident from the sequence alone.

Example 89

30 The following *C.pneumoniae* protein (PID 4377305) was expressed <SEQ ID 177; cp7305>:

35	1 MEVYSFH PAV RTSFQHRVMA ALDAWFFLGG HRLKVVS LDS CNSGWAYQEL
	51 VSISTTEKVL KLLSYLLVPI VIIALLIRCL LHSNFRIDVE KERWLKIREL
	101 GIDIESCKLP SSYVNQVSSF IWFEKDKSKR PRIDVDYHTL HSKDWVVFP I
	151 VFQKIPKTSR FS YWFQS QKET RKR DYVRNML DHVIGYL TSE GG EWILQYISK
	201 TSYQSATS LDP ERV LQY CLT DNQ ELQ GEV Q RLLNEESATK SSGDKEVLLS
	251 HVSDII CQC W WPKFLEV IQS PAFIEELV EEE VSGKLNLDFL CLEKANTLDQ
	301 ELRNSLLRAV VHHGSEGVDI KKVGAGLIY TEAIQLQIPF SRS*

The cp7305 nucleotide sequence <SEQ ID 178> is:

40	1 ATGGAAGTTT ATAGTTTCA CCCTGCGGTA AGGACTTCGT TTCAGCACCG
	51 TGTAAATGGCA GCACTAGATG CTTGGTTTT TCTAGGAGGG CACCGTTAA
	101 AAGTAGTTTC TCTAGATAGT TGTA ACTCAG GTTGGCGTA TCAAGAACTT
	151 GTGTCTATTT CAACGACAGA AAAAGTCTTG AAACTACTCT CTTACCTACT
	201 CGTACCGATT GTCATAATAG CTC TGTAAAT TCGTTGTCTT TTACATAGCA
	251 ATTTTAGGAT AGACGTAGAG AAGGAACGTT GGTTAAAAT AAGGGAGTTA
45	301 GGAATTGATA TAGAAAGCTG CAAACTCCCC AGTTCTTATG TAAACCAGGT
	351 TTCCTCGTTT ATTTGGTTG AAAAGATAA ATCCAAACGG CCACGTATTG
	401 ATGTAGATTA TCATACGCTA CATAGCAAAG ACTGGGTAGT TTTCCCTATC

5 51 TTTACCTTGT TGTGCTGCAC GAAGACGTGC TTCTGGAGAA AATTGCAAC
 101 AAACTCGTCC TATAGCAGCT GCAAATCTAC AATGGGAGAG CTATGCAGAA
 151 GCTCTTGAAAC ATTCTAAACA AGATCACAAA CCTATTGTC TTTTCTTAC
 201 AGGATCAGAC TGGTGTATGT GGTGCATAAA AATGCAAGAC CAGATTTGC
 251 AAAGCTCTGA GTTAAAGCAT TTTGGGGTG TGCATCTGCA TATGGTTGAA
 301 GTTGATTTCC CCCAAAGAA TCATCAACCT GAAGAGCAGC GCCAAAAAAA
 351 TCAAGAACTG AAAGCTCAAT ATAAAGTTAC AGGATTCCCC GAACTGGTCT
 401 TCATAGATGC AGAAGGAAAA CAGCTTGCTC GCATGGGATT TGAGCCTGGT
 451 GGTGGAGCTG CTTACGTAAG CAAGGTGAAG TCTGCTCTTA AACTACGTTA
 501 A

10 The PSORT algorithm predicts periplasmic (0.932).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 86A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 86B) and for FACS analysis.

15 The cp7248 protein was also identified in the 2D-PAGE experiment.

These experiments show that cp7248 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 87

The following *C.pneumoniae* protein (PID 4377249) was expressed <SEQ ID 173; cp7249>:

20 1 MIPSPPTPINF RDDTILETDP KPSLIMFSSK KTEIASERRK AHPTLFKVLG
 51 TIWNIVKFII SIILFLPLAL LWVLLKKTCQF FILPSSIISQ SMSKTAVALR
 101 RMTFLSHIKQ LLSLKEISAA DRVVIQYDDL VVDSLAIKIP HALPHRWILY
 151 SOGNGLMEN LFDRGDSSLH QLAKATGSNL LVFNYPGIMS SKGEAKRENL
 201 VKSYQACVRY LRDEETGPKA NQIIAFGYSL GTSVQAAALD REVTDGSDGT
 251 SWIVVKDRGP RSLADVANQI CKPIASAIIK LVGWNIDSVK PSERLRCPEI
 301 FIYNSNHDQE LISDGLFERE NCVATPFLEL PEVKTSGTKI PIPERDLLHL
 351 NPLSPNVVDR LAAVISNYLD SENRKSQQPD *

The cp7249 nucleotide sequence <SEQ ID 174> is:

30 1 ATGATCCCAT CCCCTACCCCC AATAAACTTT CGTGATGATA CGATTCTAGA
 51 GACGGATCCA AAGCCGTCTT TAATCATGTT CTCTTCAAAA AAAACAGAGA
 101 TAGCTTCTGA AAGACGGAAG GCCCCATCCCCA CCTTATTATAA AGTTCTAGGA
 151 ACCGATTGGA ATATTGTGAA GTTTATTATC TCAATCATTC TGTTCTTCC
 201 CTTAGCGTTA TTGTGGGTAC TCAAGAAAAC CTGTCAGTTT TTCATTCTCC
 251 CATCTTCTAT CATATCTCAG AGCATGTCAA AACAGCTGT GGCAATTCCG
 301 CGAATGACCT TTCTGTCCCA TATTAACCAA CTCTTAAGCC TTAAGGAAAT
 351 CTCAGCTGCC GATCGTGTGG TTATACAATA TGACGATTG GTGGTTGATA
 401 GCTTAGCTAT AAAGATACCT CATGCTCTTC CCCACAGGTG GATTCTTTAT
 451 TCTCAAGGAA ACTCTGGATT GATGGAAAAC CTGTTCGATC GGGCGATTG
 501 CTCTCTACAC CAGCTAGCCA AAGCAACCGG CTCGAATCTT CTTGTGTTCA
 551 ACTATCCTGG AATTATGTCC AGCAAAAGGAG AAGCGAACG AGAAAATCTG
 601 GTTAAATCGT ATCAGGGCATG CGTACGCTAC CTACGAGATG AAGAGACAGG
 651 TCCTAAAGCC AATCAAATCA TAGCTTTCGG ATACTCTTG GGAACTAGTG
 701 TCCAAGCTGC TGCTCTAGAT CGTGAGGTCA CTGATGGCAG TGATGGAACG
 751 TCATGGATTG TTGTAAAAGA TCGGGGCCCT CGCTCTCTAG CAGATGTCGC
 801 GAATCAAATT TGTAAGCCCA TAGCTTCCGC GATTATAAAA CTCGTTGGTT
 851 GGAACATAGA CTCTGTGAAA CCTAGCGAAA GATTGCGTTG TCCCGAAATT
 901 TTCAATTCTACA ACTCTAAATCA TGATCAAGAA CTCAATTAGCG ACGGCCTCTT
 951 CGAAAGAGAA AATTGCGTAG CAACACCTTT TCTAGAGCTT CCTGAAGTAA
 1001 AACCTCGGG GACTAAAATT CCTATACCCG AAAGGGATCT TCTCCATCTA
 1051 AATCCTCTCA GTCCAAATGT AGTAGACAGA TTAGCAGCAG TGATCTCTAA
 1101 TTATTTAGAT TCTGAAAACA GAAAGTCTCA GCAACCTGAT TAA

The PSORT algorithm predicts inner membrane (0.571).

5

```

1051 CATTGGAAA AAGAGACTGA TGCTTGATT ATTGATCAGA CCCATAATCC
1101 TGGAGGCAGT GTTTCTATC TCTATTGTT ACTATCTATG TTAACAGATC
1151 ATCCCTTAGA TACTCCTAAA CATAGAATGA TTTTCACTCA GGATGAAGTC
1201 AGCTCGGCTT TGCACTGGCA AGATCTACTA GAAGATGCTC TCACAGATGA
1251 GCAGGCAGTT GCGGTGCTAG GGGAAACTAT GGAAGGATAT TGCACTGGATA
1301 TGCATGCTGT AGCCTCTCTT CAAAACCTCT CTCAGAGTGT CCTTTCTTCC
1351 TGGGTTTCAG GTGATATTAA CCTTCAAAAA CCTATGCCCT TGCTAGGATT
1401 TGCACAGGTT CGACCTCATC CTAAACATCA ATATACTAAA CCTTTGTTTA
1451 TGGTGTAGA CGAGGATGAC TTCTCTGTG GAGATTTAGC GCCTGCAATT
1501 TTGAAGGATA ATGGCCGCGC TACTCTCATC GGAAAGCCA CAGCAGGAGC
1551 TGGAGGTTT GTATTCCAAAG TCACCTTCCC TAACCGTTCT GGAATTAAAG
1601 GTCTTCTTT AACAGGATCT TTAGCTGTTA GGAAAGATGG TGAGTTTATT
1651 GAAAACCTAG GAGTGGCTCC TCATATTGAT TTAGGATTAA CCTCCAGGGA
1701 TTGCAAAC TCCAGGTTA CTGATTACGT TGAGGCAGTG AAAACTATAG
1751 TTTAACCTTC TTGCTCTGAG AACGCTAAGA AGAGTGAAGA GCAGACTTCT
1801 CCGCAAGAGA CGCCTGAAGT TATTCGAGTC TCTTATCCCA CAACGACTTC
1851 TGCTTCGTAA

```

The PSORT algorithm predicts periplasmic space (0.2497).

20 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 90A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 90B) and for FACS analysis.

These experiments show that cp7347 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 91

25 The following *C.pneumoniae* protein (PID 4377353) was expressed <SEQ ID 181; cp7353>:

30

```

1 MNMPVPSAVP SANITLKEDS STVSTASGIL KTATGEVLVS CTALEGSSST
51 DALISLALGQ IILATQQELL LQSTNVHQLL FLPPEVVELE IQVV DLLVQL
101 EHAETITSEP QETQTQSRSE QTLPPQQSSK QSALSPRSLK PEISDSKQQQ
151 ALQTPKDSAV RKHSEAPSPE TQARASLSQA SSSSQRSLPP QESAPERPLL
201 EQQKASSFSP LSQFSAEKQK EALTTSKSHE LYKERDQDRQ QREQHDRKHD
251 QEEDAEEKKK KKKRGLGVEA VAEEPGENLD IAALIFSDQM RPPAEETSKK
301 ETTFKKLPS PMSVFSRFIP SKNPLSVGSS IHGPIQTPKV ENVFLRFMKL
351 MARILGQAEA EANEELYMRVK QRTDDVDTLT VLISKINNEK KDIDWSENEE
401 MKALLNRAKE IGVTIIDKEKY TWTEEEKRLL KENVQMRKEN MEKITQMERT
451 DMQRHLQEIS QCHQARSNVL KLLKELMDTF IYNLRP*

```

The cp7353 nucleotide sequence <SEQ ID 182> is:

40

```

1 ATGAATATGC CTGTTCCCTTC TGCAGTTCCC TCTGCCAAATA TAACTCTAAA
51 AGAACAGACAGC TCAACAGTTT CCACAGCCTC TGGAAATATTA AAGACTGCAA
101 CAGGTGAAGT CTTAGCTCTCT TGTACAGCGC TAGAAGGAAG CTCTTCTACA
151 GATGCTTTAA TTAGCTTAGC TTTAGGACAA ATCATTCTTG CGACCCAACA
201 AGAACCTGCTC TTACAAAGCA CAAATGTTCA TCAACTCCTC TTCCCTCCCTC
251 CTGAAGTTGT AGAATTAGAA ATCCAAGTTG TTGACTTGCT AGTGAATTG
301 GAAACATGCGAG AGACAATCAC AAGTGAACCA CAAGAAACAC AAACGCAAAG
351 TAGGAGTGTAG CAGACCCCTCC CTCAACAAAG CAGCAGTAAA CAATCTGCTC
401 TCTCCCCACAG CTCCTTAAAAA CCTGAAATTCTT CTGATTCTAA ACAACAGCAA
451 GCTCTTCAAA CACCAAAAGA CTC TGCTGTAA AGAAAACACA CGGAAGCACC
501 GTCACCTGAG ACACAAAGCTC GCGCTTCCCTT ATCTCAGGCA AGCTCAAGTT
551 CTCAGAGATC CTTACCTCCG CAAGAAAGTG CGCCAGAAAG AACACTATTA
601 GAAACAACAAA AAGCAAGCTC CTTCTCTCCT CTATCCCAGT TCTCTGCAGA
651 GAAACAAAAA GAGGCCCTGA CGACCTCAAA ATCTCATGAA CTCTATAAAG
701 AACCGCAGTC AGATCGCCAA CAAAGAGAGC AGCACGACAG AAAGCACGAT
751 CAGGAAGAAG ACGCTGAATC TAAAAAGAAA AAGAAGAAAC GTGGTCTCGG
801 TGTAGAGGCA GTCGCTGAGG AACCCGGAGA AAATCTAGAT ATTGCCGCTT
851 TAATCTTCTC AGATCAAATG CGACCTCCCTG CTGAAGAAC TTCTAAAAAA
901 GAAACGACAT TCAAAAGAA GCTACCTCT CCAATGTCTG TGTTTAGCAG
951 ATTCACTCCCT AGTAAGAATC CGTATCTGT AGGCTCTCA ATACACGGGC
1001 CTATACAAAC TCCAAAAGTA GAAAATGTGT TCTTAAGGTT CATGAAGCTC

```

5 451 GTTTTCAGA AAATTCCAAA GACCTCGCGT TTCAGTTATT GGTTCTCACA
 501 AAAAGAAACA AGGAAGAGGG ATTATGTGAG AAATATGCTG GACCACGTCA
 551 TTGGTTATCT AACGTCAGAA GGTGGGAGT GGTTGCAGTA TATATCGAAA
 601 ACCTCTTATC AAAGCGCTAC TTCCCTGGAT CCTGAAAGAG TTCTTCATA
 651 TTGCTTAACT GATAACCAGG AGCTCCAGGG AGAAGTGCAA CGTTTGCTTA
 701 ATGAGGAGAG TGCGACCAAA AGCTCTGGGG ATAAGGAAGT TTTGTTAAGT
 751 CATGTATCTG ACATTATTTG CCAGTCITGG TGGCCAAGT TTCTTGAAAGT
 801 TATACAATCT CCGGCCTTA TTGAAGAAATT AGTAGAAGAA GTGAGTGGTA
 851 AACTTAATTT AGATTTTTA TGCCTAGAAA AGGCTAATAC ATTAGATCAG
 901 GAGTTGAGAA ACAGTCTCT AAGAGCAGTC GTACACCACG GTTCTGAAGG
 951 AGTTGATATT AAGAAAGTTG GTGCCGGCCT CAATTATTAT ACGGAAGCTA
 10 1001 TTCAATTACA GATTCCCTTC TCAAGGAGTT AA

The PSORT algorithm predicts inner membrane (0.508).

15 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 89A) and also as a double GST/his fusion. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 89B) and for FACS analysis.

These experiments show that cp7305 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 90

20 The following *C.pneumoniae* protein (PID 4377347) was expressed <SEQ ID 179; cp7347>:

25 1 MKKGKLGATV FGLLFTSSVA GFSKDLTKDN AYQDLNVIEH LISLKYAPLP
 51 WKELLFGWDL SQQTQQARLQ LVLEEKPTTN YCQKVLSNYV RSLNDYHAGI
 101 TFYRTESEAYI PYVLKLSEDG HVFVVVDVQTS QGDIYLGDEI LEVDGMGIRE
 151 AIESLRFGRG SATDYSAAVR SLTSRSAAFG DAVPSGIAML KLRRPSSLIR
 201 STPVRWRWRYTP EHIGDFSLVA PLIPEHKPQL PTQSCVLFRS GVNSQSSSSS
 251 LFSSYMPVYF WEELRVQNQKQ RFDSNHHIGS RNGFLPTFGP ILWEQDKGPY
 301 RSYIFKAKDS QGNPHRJIGFL RISSYVWTDL EGLEEDHKDS PWELFGEIID
 351 HLEKETDALI IDQTHNPGGS VFYLYSLLSM LTDHPLDTPK HRMIIFTQDEV
 401 SSALHWQDLL EDVFTDEQAV AVLGETMEGY CMDFMHAVASL QNFSQSVLSS
 451 WVSGDINLSK PMPLLGFQAQV RPHPKHQYTK PLFMLIDEDD PSCGDLAPAI
 501 LKDNGRATLI GKPTAGAGGF VFQVTFPNRS GIKGLSLTGS LAVRKDGIFI
 551 ENLGVAPHID LGFTSRDLQT SRFTDYVEAV KTIVLTSLSE NAKKSEEQTS
 601 PQETPEVIRV SYPTTTSAS*

A predicted signal peptide is highlighted.

35 The cp7347 nucleotide sequence <SEQ ID 180> is:

40 1 ATGAAAAAAAG GGAAATTAGG AGCCATAGTT TTTGGCCTTC TATTTACAAG
 51 TAGTGTGCT GGTTTTCTA AGGATTGAC TAAAGACAAC GCTTATCAAG
 101 ATTTAAATGT CATAGAGCAT TTAATATCGT TAAAATATGC TCCTTTACCA
 151 TCGAAGGAAC TATTATTG TGCGGATTAA TCTCAGCAA CACAGCAAGC
 201 TCGCTTGCAA CTGGTCTTAG AAGAAAAACC AACAAACCAAC TACTGCCAGA
 251 AGGTACTCTC TAACTACGTG AGATCATTAA ACGATTATCA TGCAGGGATT
 301 ACGTTTTATC GTACTGAAAG TGCGTATATC CCTTACGTAT TGAAGTTAAC
 351 TGAAGATGGT CATGTCTTTG TAGTCGACGT ACAGACTAGC CAAGGGGATA
 401 TTTACTTAGG GGATGAAATC CTTGAAGTAG ATGGAATGGG GATTCTGAG
 451 GCTATCGAAA GCCTTCGCTT TGGACGAGGG AGTGCCACAG ACTATTCTGC
 501 TGCAGTTCGT TCCTTGACAT CGCGTTCCGC CGCTTTGGA GATGCGGTTC
 551 CTTCAGGAAT TGCCATGTTG AAACCTCGCC GACCCAGTGG TTTGATCCGT
 601 TCGACACCGG TCCGTTGGCG TTATACTCCA GAGCATATCG GAGATTTTC
 651 TTTAGTTGCT CCTTTGATTC CTGAAACATAA ACCTCAATTAA CCTACACAAA
 701 GTTGTGTGCT ATTCCGTTCC GGGGTAAATT CACAGTCCTTC TAGTAGCTCT
 751 TTATTCAGTT CCTACATGGT GCCTTATTC TGGGAAGAAT TGCGGGTTCA
 801 AAATAAGCAG CGTTTTGACA GTAATCACCA TATAGGGAGC CGTAATGGAT
 851 TTTTACCTAC GTTGGTCC ATTCTTTGGG AACAAAGACAA GGGGCCCTAT
 901 CGTTCCCTATA TCTTTAAAGC AAAAGATTCT CAGGGCAATC CCCATCGCAT
 951 AGGATTTTTA AGAATTCTT CTTATGTTG GACTGATTAA GAAGGACTTG
 55 1001 AAGAGGATCA TAAGGATAGT CCTTGGGAGC TCTTGGAGA GATCATCGAT

Example 93

The following *C.pneumoniae* protein (PID 4376424) was expressed <SEQ ID 185; cp6424>:

```

5   1 MMHNIVVLSE EPGRSAFLGR TAFFPNKYPI AQGGVGIPST IGNLFTIWC
  51 51 FYFYRAATPQ SDHPDGCIFI LLERLKELGA GFFYCDLRES NTTGFTLFFE
  101 101 GSNKGVLKNH LFIRDE*

```

The cp6424 nucleotide sequence <SEQ ID 186> is:

```

10  1 ATGATGCCACA ATATTGTGTC TCTTAGTGAG GAACCTGGAC GAAGCGCTTT
  51 51 TCTTGGTAGG ACGGCATTT TCCCTAAATA GTATCCAATA GCTCAGGGTG
  101 101 GTGTTGGAAT ACCATCTACA ATAGGCAATC TCTTTACTAT ATGGTACTGT
  151 151 TTCTATTTTT ATAGAGCTGC AACTCCACAA TCTGATCATC CTGACGGATG
  201 201 TGCGTTTATT CTACTAGAAA GGCTTAAGGA GCTCGGTGCA GGGTTCTTTT
  251 251 ATTGTGATCT TCGTGAGTCC AATACCACTG GCTTTACTCT TTTTTTGAA
  301 301 GGCTCCAATA AAGGTGTGTT AAAGAACAC TTGTTTATTA GAGATGAGTA
  351 351 A

```

15 The PSORT algorithm predicts cytoplasm (0.2502).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 93A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figure 93B) and for FACS analyses (Figure 93C; GST-fusion).

20 These experiments show that cp6424 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 94

The following *C.pneumoniae* protein (PID 4376449) was expressed <SEQ ID 187; cp6449>:

```

25  1 VASETYPSQI LHAQREVRDA YFNQADCHPA RANQILEAKK ICLLDVYHTN
  51 51 HYSVFTFCVD NYPNLRFTFV SSKNNEMLN SNPLDNVLVE AMVRRTHARN
  101 101 LLAACKIRNI EVPRVVGLDL RSGILISKLE LKQPQFQS LT EDFVNHSNTNQ
  151 151 EEARVHQKHV LLISLILLCK QAVLESFQEKRSS*

```

The cp6449 nucleotide sequence <SEQ ID 188> is:

```

30  1 GTGGCGTCTG AAACGTATCC TTCTCAGATA TTGCACGCTC AGAGGGAAGT
  51 51 ACGTGATGCC TATTTAAC T AAGCGGATTG CCATCCTGCT CGGGCTAAC
  101 101 AGATTCTCGA GGCTAAGAAA ATCTGTTTAT TAGATGTTA TCATACTAA
  151 151 CATTATTCCG TATTTACTTT TTGTGTAGAT AATTATCCGA ATCTCCGCTT
  201 201 TACATTTGTA TCTTCAAAAA ACAATGAGAT GAATGGCTTA TCTAATCCTC
  251 251 TAGATAATGT TCTTGTAGAG GCTATGGTAC GTAGAACACA TGCAAGAAC
  301 301 CTACTTGCAG CGTGTAAAT TCGAAATATT GAGGTTCCAA GGGTTGTTGG
  351 351 GCTTGACCTA AGATCTGGGA TACTCATTTG GAAACTAGAA TTGAAGCAAC
  401 401 CTCAGTTCCA AAGTTAAACA GAAGACTTCG TAAATCATTC CACAAATCAG
  451 451 GAAAGAAGCTC GCGTCCATCA AAAGCATGTG TTGCTAATT CTTTAATT
  501 501 ACTTGCAAG CAGGCCGTT TGGAAATCATT CCAGGAAAAA AAGCGATCCT
  551 551 CTAA

```

40 The PSORT algorithm predicts inner membrane (0.2084).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 94A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figure 94B) and for FACS analyses (Figure 94C; GST-fusion).

45 These experiments show that cp6449 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

5
1051 ATGGCAAGAA TCTTAGGCCA AGCCGAAGCC GAAGCTAATG AACTCTACAT
1101 GCGAGTCAAA CAACGTACCG ATGATGTAGA CACACTCACA GTCCTTATCT
1151 CTAAGATCAA TAATGAAAAG AAAGACATTG ATTGGAGTGA AAATGAAGAG
1201 ATGAAAGCTC TTTTAAATCG AGCTAAAGAG ATTGGAGTCA CTATAGACAA
1251 AGAAAAATAT ACTTGACAG AAGAGGAAAA AAGACTTCTA AAAGAGAATG
1301 TCCAAATGCG CAAAGAGAAT ATGGAGAAAA TCACTCAAAT GGAAAGGACG
1351 GACATGCAAA GGCACCTCCA AGAGATTCT CAATGTCATC AAGCGCGCTC
1401 TAATGTATTG AAGTATTG AAGAACTTAT GGACACCTTC ATTTACAACC
1451 TACGCCCTA A

10 The PSORT algorithm predicts cytoplasm (0.1308).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 91A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 91B) and for FACS analysis.

15 These experiments show that cp7353 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 92

The following *C.pneumoniae* protein (PID 4377408) was expressed <SEQ ID 183; cp7408>:

20
1 MLKIQKKRMC VSVVITVGAI VGFFNSADAA PKKKKIPIQI LYSFTKVSSY
51 LKNEDASTIF CVDVDRGLLQ HRYLGSPGWQ ETRRRQLFKS LENQSYGNER
101 LGEETLAIDI FRNKECLESE IPEQMEAILA NSSALVLGIS SFGITGIPAT
151 LHSLLRQNLS FQKRSIASES FLLKIDSAPS DASVFYKGVL FRGETAIVDA
201 LSQLFAQQLDL SPKKIIFLGE DPEVVAQVGS ACIGWGMNFL GLVYYPAQES
251 LFSYVHPYST ATELQEAQQL QVISDEVAQL TLNALPKMN*

The cp7408 nucleotide sequence <SEQ ID 184> is:

25
1 ATGTTGAAAA TCCAGAAAAA AAGAATGTGT GTCAGCGTAG TCATCACGGT
51 AGGCGCCATA GTGGGGTTTT TCAATTCTGC AGACGCAGCA CCAAAGAAAA
101 AGAAGATCCC TATACAGATT CTCTACTCCT TTACTAAAGT CTCTTCCTAT
151 TTAAAAAACG AAGACGCAAG TACTATATTT TGCGTCGATG TGGATCGTGG
201 ACTTCTCCAG CATCGGTATT TAGGTAGTCC AGGATGGCAG GAAACCAGAC
251 GTCGGCAGTT ATTTAAATCC TTAGAAAATC AATCATACGG CAACGAACGT
30 301 TTAGGAGAAG AAACTCTTGC TATTGATATT TTCAGGAACA AAGAGTGCTT
35 351 GGAGAGCGAG ATCCCAGAGC AGATGGAAGC TATCCTTGCA AATTCTCGG
40 401 CCTTGGTCTT AGGCATCTCT TCTTTGGGA TCACAGGAAT TCCTGCGACT
451 TTGCGATGTT TGCTTCGACA GAATCTATCT TTCCAAAAAC GCTCTATAGC
501 ATCGGAGAGC TTCTTTTAA AGATCGATAG TGCCCCCTCA GATGCTCTG
551 TTTTTTATAA AGGCGTGCTT TTCCCGGGAG AGACTGCGAT CGTGGATGCG
601 TTAAGCCAAT TATTTGCCA GCTCGATCTT TCTCCTAAAA AAATTATCTT
651 TCTAGGAGAA GACCCCTGAGG TCGTTCAAGC TGTTGGGTCT GCTTGTATAG
701 GTTGGGGCAT GAACTTTTA GGCCTGGTAT ACTATCCTGC TCAAGAAAGC
751 CTTTTTCTT ATGTTCATCC TTACTCTACA GCAACGGAGC TCCAAGAAGC
801 ACAGGGTTTA CAAGTAATT CAGATGAAGT CGCACAGCTT ACTTTAAACG
851 CTCTTCCGAA AATGAATTAA

The PSORT algorithm predicts inner membrane (0.123).

45 The protein was expressed in *E.coli* and purified as a his-tag product (Figure 92A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 92B) and for FACS analysis.

These experiments show that cp7408 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

These experiments show that cp6506 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 97

The following *C.pneumoniae* protein (PID 4376882) was expressed <SEQ ID 193; cp6882>:

```

5      1  MSLLNLPPSQ DSASEDSTSQ SQIFDPIRNR ELVSTPEEKV RQRLLSFLMH
      51 KLNYPKKLII IEKELKTLFP LLMRKGTLLIP KRRPDILIT PPTYTDAQGN
     101 THNLGDPKPL LLIECKALAV NQNALKQLLS YNYSIGATCI AMAGKHSQVS
     151 ALFPNPKTQTL DFYPGLPEYS QLLNYFISLN L*

```

The cp6882 nucleotide sequence <SEQ ID 194> is:

```

10     1  ATGTCCTTAT TGAACCTTCC CTCAAGCCAG GATTCTGCAT CTGAGGACTC
      51 CACATCGCAA TCTCAAATCT TCGATCCCAT TAGAAATCGG GAGTTAGTTT
     101 CTAECTCCGA AGAAAAAGTC CGCCAAAGGT TGCTCTCCCT CCTAATGCAT
     151 AAGCTGAACT ACCCTAAAGAA ACTCATCATC ATAGAAAAG AACTCAAAAC
     201 TCTTTTTCCCT CTGCTTATGC GTAAAGGAAC CCTAATCCCA AAACGCCGCC
     251 CAGATATTCT CATCATCACT CCCCCACAT ACACAGACGC ACAGGGAAAC
     301 ACTCACAACC TAGGCGACCC AAAACCCCTG CTACTTATCG AATGTAAGGC
     351 CTTAGCCGTA AACCAAAATG CACTCAAACA ACTCCTTAGC TATAACTACT
     401 CTATCGGAGC CACCTGCATT GCTATGGCAG GGAAACACTC TCAAGTGTCA
     451 GCTCTCTTCA ATCCAAAAAC ACAAACTCTT GATTTTATC CTGGCCTCCC
     501 AGAGTATTCC CAACTCCTAA ACTACTTTAT TTCTTTAAC TTATAG

```

The PSORT algorithm predicts cytoplasm (0.362).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 97A). The protein was used to immunise mice, whose sera were used in a Western blot (Figure 97B) and for FACS analysis (Figure 97C).

25 These experiments show that cp6882 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 98

The following *C.pneumoniae* protein (PID 4376979) was expressed <SEQ ID 195; cp6979>:

```

30     1  MSVNPSGN SK NDLWITGAHD QHPDVKEESGV TSANLGSHRV TASGGRQGLL
      51 ARIKEAVTGF FSRMSFRSG APRGSQQPSA PSADTVRSPL PGGDARATEG
     101 AGRNLLIKKGY QPGMKVTIPQ VPGGQAQRSS GSTTLKPTRP APPPKTGGT
     151 NAKRPATHGK GPAPQPQPKT GTNAKRAATH GKGPAPQPPK GILKQPGQSG
     201 TSGKKRVSWS DED*

```

The cp6979 nucleotide sequence <SEQ ID 196> is:

```

35     1  ATGTCGTGTTA ATCCATCAGG AAATTCCAAG AACGATCTCT GGATTACGGG
      51 AGCTCATGAT CAGCATCCCG ATGTTAAAGA ATCCGGGGTT ACAAGTGCTA
     101 ACCTAGGAAG TCATAGAGTG ACTGCCTCAG GAGGACGCCA AGGGTTATTA
     151 GCACGAATCA AAGAACGAGT AACCGGGTTT TTTAGTCGGA TGAGCTTCTT
     201 CAGATCGGGA GCTCCAAGAG GTAGCCAACA ACCCTCTGCT CCATCTGCAG
     251 ATACTGTACG TAGCCGTTG CGGGGAGGGG ATGCTCGCCG TACCGAGGG
     301 GCTGGTAGGA ACTTAATTAA AAAAGGGTAC CAACCCAGGGA TGAAAGTCAC
     351 TATCCCACAG GTTCCTGGAG GAGGGGCCCA ACGTTCATCA GGTAGCACGA
     401 CACTAAAGCC TACCGTCCG GCACCCCCAC CTCCCTAAAAC GGGTGGAACT
     451 AATGCAAAAC GTCCGGCAAC GCACGGGAAG GGTCCAGCAC CCCAGCCTCC
     501 TAAAACAGGT GGGACCAATG CTAAGCGCGC AGCAACGCAT GGGAAAGGTC
     551 CAGCACCTCA ACCTCTTAAG GGCATTTGA AACAGCCTGG GCAGTCTGGG
     601 ACTTCAGGAA AGAAGCGTGT CAGCTGGTCT GACGAAGATT AA

```

The PSORT algorithm predicts cytoplasm (0.360).

Example 95

The following *C.pneumoniae* protein (PID 4376495) was expressed <SEQ ID 189; cp6495>:

MRELNAFELTQPEEYRNRWVLMPCLKCRFCRTQHAKVWSYRCVHEASLYEKNCFLTLTYDDKHL PQYGSILVKLHLQLFLKR
LRKMISPHKIRYFECGAYGTLQRPHYHLLS

- 5 The cp6495 nucleotide sequence <SEQ ID 190> is:

TTGCGAGAATTAAATGCTTTGAATTAACTCAACCTGAAGAGTATCGAAACCGTTGGGTGGATGCCCTGTCTTAAGTGT
CGTTTTGTAGAACGCAACATGCAAAGCTGGTCTATCGTTGTCCATGAAGCTTCCTTGATGAGAAAAATTGTTT
CTTACTTTGACTTATGATGATAAGCATTACCTCAGTATGGTTGTGGTAAAGCTGCATTACAGCTGTTCTTAAGAGA
TTAAGAAAGATGATTCTCCTCATAAAATTGTTATTTGAATGTGGTGCATGGAACCAAATTACAAGACCTCATTAT
10 CATCTACTTTTATCATGA

The PSORT algorithm predicts cytoplasmic (0.280).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 95A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 95B) and for FACS analysis (Figure 95C).

- 15 These experiments show that cp6495 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 96

The following *C.pneumoniae* protein (PID 4376506) was expressed <SEQ ID 191; cp6506>:

20 1 MRRFLFLILS SLPLVAFSAD NFTILEEKQS PLSRVSIIFA LPGVTPVSFD
51 51 GNCPIPWFSH SKKTLEGQRI YYSGDSFGKY FVVSALWPNK VSSAVVACNM
101 101 ILKHRVDLIL IIGSCYSRSQ DSRGFSVLVS KGYINYDADV RPFFERFEIP
151 151 DIKKSVFATS EVHREAILRG GEEFISTHKQ EIEELLKTHG YLKSTTKTEH
201 201 TLMEGLVATG ESFAMSRNYF LSLQKLYPEI HGFDSVSGAV SQVCYEYSIP
251 251 CLGVNILLPH PLESRSNEDW KHLQSEASKI YMDTLLKSVL KELCSSH*

- 25 The cp6506 nucleotide sequence <SEQ ID 192> is:

1 1 ATGCGTCGTT TTCTGTTCT TATTCTTAGC TCTCTTCCTT TGGTCGCATT
51 51 CTCTGCTGAT AATTTCACTA TTCTAGAAGA AAAACAGAGT CCTTTAACGTC
101 101 GTGTAAGTAT TATTTTGCT TTACCTGGGG TTACTCCCGT TTCTTTGAT
151 151 GGTAAATTGTC CTATTCCTTG GTTTCTCAT AGTAAAAAGA CTCTAGAGGG
201 201 ACAGAGAATT TATTACTCTG GCGACTCCTT TGGGAAATAC TTTGTAGTTT
251 251 CTGCTCTTTG GCCTAATAAA GTTCTTCAG CTGTTGTGGC TTGTAATATG
30 301 ATTCTTAAAC ATCGAGTGGA TCTTATTCTA ATTATAGGCT CGTGTACTC
351 351 TAGGTCTCAA GATAGCCGTT TTGGCAGCGT CTTAGTTCT AAAGGCTACA
401 401 TTAATTATGA TGCAGATGTG AGGCCTTTCT TTGAAAGATT TGAGATTCCA
35 451 GACATTAAAA AGAGTGTGTT TGCAACCAGT GAGGTTCATC GGGAGGCAAT
501 501 TCTTCGTGGA GGCAGAGT TTATTTCTAC CCATAAACAA GAAATCGAAG
551 551 AGCTTTGAA GACTCATGGG TATTTGAAAT CAACAACCAA AACGGAGCAC
601 601 ACCTTAATGG AAGGTTTGGT TGCTACAGGC GAGTCCTTCG CGATGTCGCG
651 651 AAACTATTCTT CTTCCCTTAC AAAAATTGTA TCCAGAGATT CATGGTTTTG
40 701 701 ATAGTGTCA CGGCGCTGTT TCTCAGGTAT GCTATGAATA TAGCATTCCCT
751 751 TGTGTTAGGTG TGAATATCCT TCTCCCTCAT CCTTTAGAAT CACGGAGTAA
801 801 CGAGGATTGG AAGCATCTTC AAAGTGAGGC AAGTAAAATT TATATGGATA
851 851 CCTTGCTCAA GAGTGTATTA AAAGAACTCT GTTCTTCTCA TTAA

The PSORT algorithm predicts periplasmic space (0.571).

- 45 The protein was expressed in *E.coli* and purified as his-tag (Figure 96A) and GST-fusion (Figure 96B) products. The GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 96C) and for FACS analysis (Figure 96D).

251 AATGTTATAAC CCGATTGAA GATGGCACAA TTTTTTATGA ATGCGATTAG

The PSORT algorithm predicts inner membrane (0.143).

The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 100A) and a his-tag product. The proteins were used to immunise mice, whose sera were used in a Western blot (Figure 5 100B) and for FACS analysis (Figure 100C).

These experiments show that cp7355 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 101

The following *C.pneumoniae* protein (PID 4377380) was expressed <SEQ ID 201; cp7380>:

10	1 VHYCERTLDP KYILKIALKL RQSLSLFFQN SQSLQRAYST PYSYYRIILQ
	51 KENKEKQALA RHKCISILEF FKNNLLFVHLL SLSKNQREGC STDMAVVSTP
	101 FFNRNLWYRL LSSRFSLWKS YCPRFFLDYL EAFFGLLSDFL DHQAVIKFFE
	151 LETHFSYYPV SGFVAPHQYL SLLQDRYFPPI ASVMRTLDKD NFSLTPDLIH
15	201 DLLGHVPWLL HPSFSEFFIN MGRLFTKVIE KVQALPSKKQ RIQTLQSNLI
	251 AIVRCFWFTV ESGLIENHEG RKAYGAVLIS SPQELGHAFI DNVRVLPLEL
	301 DQIIRLPFNT STPQETLFSI RHFDELVELT SKLEWMLDQG LLESIPLYNQ
	351 EKYLSGFEVL CQ*

The cp7380 nucleotide sequence <SEQ ID 202> is:

20	1 GTGCACTACT CGCAGAGAAC CCTGGACCCA AAGTATATTTC TGAAGATTGC
	51 TCTAAAGCTG AGACAATCAC TTTCCCTGTT CTTCCAGAAC AGCCAATCAC
	101 TCCAACGTGC ATACTCGACC CCATATTCCCT ACTACCGAAT CATTCTACAA
	151 AAGGAAAATA AAGAGAAGCA AGCTTTAGCT CGACACAAAT GCATTTCTAT
	201 TTTAGAATTTC TTCAAAAATCT TACTCTTTGT TCATCTCTG TCATTATCAA
25	251 AGAATCAAAG GGAAGGTTGC TCCACTGATA TGGCTGTTGT AAGCACTCCC
	301 TTTTTTAATC GGAATTATATG GTATCGACTC CTTTCCCTCAC GGTTTTCTCT
	351 ATGGAAAAGC TATTGTCCAA GATTTTTCT TGATTACTTA GAAGCTTTCG
	401 GTCTCCCTTC TGATTTCTTA GACCATCAAG CAGTCATTAA ATTCTTCGAA
	451 TTAGAAACAC ATTTTCCTA TTATCCCGTT TCAGGATTG TAGCTCCCCA
30	501 TCAAATCTTG TCTCTGTTGC AGGACCGTTA CTTTCCCAT GCCTCTGTAA
	551 TCGGAACTCT CGATAAAGAT AATTCTCCT TAACTCCTGA TCTCATCCAT
	601 GACCTTTTAG GGCACGTGCC TTGGCTTCTA CATCCCTCAT TTTCTGAATT
	651 TTCATCAAAC ATGGGAAGAC TCTTCACTAA AGTCATAGAA AAAGTACAAG
	701 CTCTTCCCTAG TAAAAAACAA CGCATACAAA CCCTACAAAG CAATCTGATC
35	751 GCTATTGTC GCTGCTTTG GTTACTGTT GAAAGCGGAC TTATTGAAAA
	801 CCATGAAGGA AGAAAAGCAT ATGGAGCCGT TCTTATCAGT TCTCCTCAGG
	851 AACTTGGACA CGCTTTCATT GATAACGTAC GTGTTCTCCC TTTAGAATTG
	901 GATCAGATTA TTCTGCTTCC CTTCATACA TCAAACCTCAC AAGAGACTTT
	951 ATTTTCAATA AGACATTTG ATGAACTGGT AGAAACTCACT TCAAATTAG
40	1001 AATGGATGCT CGACCAAGGT CTGTTAGAAT CAATTCCCT TTACAATCAA
	1051 GAGAAATATC TTTCTGGTT TGAGGTACTT TGCCAATGA

The PSORT algorithm predicts inner membrane (0.1362).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 101A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 101B) and for FACS analysis (Figure 101C).

45 These experiments show that cp7380 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 102

The following *C.pneumoniae* protein (PID 4376904) was expressed <SEQ ID 203; cp6904>:

-133-

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 98A). The GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 98B) and for FACS analysis (Figure 98C).

These experiments show that cp6979 is a surface-exposed and immunoaccessible protein, and that it
5 is a useful immunogen. These properties are not evident from the sequence alone.

Example 99

The following *C.pneumoniae* protein (PID 4377028) was expressed <SEQ ID 197; cp7028>:

```

10      1 MLLGFLCDCP CASWQCAA VA NCYDSVFMSR PEHKPNIP YI TKATRRGLRM
      51 KTLAYLASLK DARQLAYDFL KDPGSLARLA KALIAPKEAL QEGNLFFYGC
      101 SNIEDILEEM RRPHRILLG FSYCQKPKAC PEGRFNDACR YDPSHPTCAS
      151 CSIGTMMRLN ARRYTTVIIP TFIDIAKHLH TLKKRYPGYQ ILFAVTACEL
      201 SLKMFQDYAS VMNLKGVGIR LTGRICNTFK AFKLAERGVK PGVTILEEDG
      251 FEVLAIRILTE YSSAPFPRDF CEIH*

```

The cp7028 nucleotide sequence <SEQ ID 198> is:

```

15      1 ATGCTTCTAG GGTTTTGTG TGACTGCC C TGTGCTTCGT GGCAGTGTGC
      51 GGCCTTGCT AATTGTTATG ATTCCGTATT TATGCTAGA CCAGAGCACA
      101 AACCTAATAT TCCTTATATT ACTAAAGCTA CAAGACGGG TCTGCGTATG
      151 AAGACGCTTG CTTATCTGGC CTCTTTAAA GATGCTAGAC AGCTTGCCTA
      201 TGATTTCTG AAAGATCCTG GTTCTTTAGC TCGGTTAGCT AAGGCTTGA
      251 TAGCTCCTAA GGAGGCCCTA CAGGAGGGCA ACCTATTTT TTATGGCTGT
      301 AGTAATATTG AGGATATTT AGAGGAGATG CGTCGTCCTC ATAGAATCCT
      351 TTTGTTAGGA TTTTCTTATT GTCAAAAGCC TAAGGCATGT CCTGAAGGGC
      401 GTTTCAATGA TGCTTGTGG TATGATCCTT CACATCCTAC ATGTGCCTCA
      451 TGTTCTATAG GGACCATGAT GCGGCTGAAT GCTCGTAGAT ACACTACTGT
      501 GATCATCCCT ACATTTATAG ATATCGCAA ACATTTACAC ACTTTAAAAAA
      551 AGCGCTACCC TGGATATCAA ATTCTCTTTG CAGTTACTGC TTGTGAACCTT
      601 TCCTTAAAAA TGTTGGAGA TTATGCCTCC GAAATGAACT TAAAGGGTGT
      651 GGGCATCAGA CTCACAGGAC GTATTTGCAA TACATTTAAG GCATTAAAT
      701 TAGCTGAGCG AGGAGTCAA CCAGGAGTCA CTATCCTAGA AGAAGATGGC
      751 TTTGAGGTAT TAGCAAGGAT TCTTACAGAA TACAGTAGCG CTCCTTTCCC
      801 TAGAGACTTT TGTGAGATCC ATTAG

```

The PSORT algorithm predicts cytoplasm (0.1453).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 99A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 35 99B) and for FACS analysis (Figure 99C).

These experiments show that cp7028 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 100

The following *C.pneumoniae* protein (PID 4377355) was expressed <SEQ ID 199; cp7355>:

```

40      1 MKVVTLSII FFATYCASEL SAVTVVAVPL SEAPGKIQVR PVVGLQFQEE
      51 QGSVPYSFYY PYDYGYYYPE TYGYTKNTGQ ESRECYTRFE DGTIFYECD*

```

The cp7355 nucleotide sequence <SEQ ID 200> is:

```

45      1 ATGAAGAAAG TCGTAACACT ATCCATTATA TTTTTCGCAA CGTATTGTGC
      51 ATCAGAGCTT AGTGTGTAA CTGTAGTGGC TGTGCTTAA TCAGAGGCTC
      101 CAGGGAAAGAT TCAAGTCGT CCCGTCGTG GTCTGCAATT TCAAGAAGAA
      151 CAGGGTTCTG TGCCCTATAG TTTTTATTAT CCTTATGACT ATGGGTATTA
      201 CTATCCAGAG ACTTATGGCT ATACTAAAAA TACAGGTCAA GAAAGTCGCG

```

-136-

```

1  LNFAKIDHNH LYLTCGLDG VACPILSTDC LPNYSEKASH EVLVYSKFRG
51  ISGEPESR LAT SGNDTYY SIV SLPIGLRYEV TSPSGRHDFN IDMHVAPKIG
101  AVLSHGTREA KEIPGSSKDY AFFSLTARES LMISEKLAMT FQVSEVIQNC
151  YSQCKVTKT NLKEQYRHL S HNTGFELSVK SAF*

```

- 5 The cp7387 nucleotide sequence <SEQ ID 208> is:

```

1  TTGAATTTG CAAAGATG A TCACAATCAT CTCTACCTTA CATGTTGGG
51  AGATCTTGGT GTAGCTTGC C CTATACTTT TACAGATTGT CTACCTAATT
101  ATAGCGAGAA AGCATCTC AT GAGGTTCTT TTTATAGTAA ATTTAGATGC
151  ATTTCTGGAG AGCCATCTCG ACTTGCAACT TCAGGAAATG ACACATATTA
10  201  TTCTATAGTA AGTTTACCTA TAGGACTCCG TTACGAAGTG ACTTCACCCT
251  CAGGACGTCA TGATTTCAAT ATTGATATGC ATGTAGCTCC AAAGATAGGT
301  GCAGTACTCT CTCATGGAAC ACGAGAGGCT AAAGAGATCC CAGGATCTC
351  AAAAGACTAT GCATTTTTA GCTTGACTGC TAGAGAAAGT TTAATGATTT
401  CTGAAAAGCT TGCGATGACT TTCCAAGTTA GCGAAGTTAT TCAGAATTGT
451  TATTCACAAT GTACTAAAGT AACGAAAAC AATTAAAAG AACAGTATAG
501  GCACTTATCC CACAATACAG GGTTGAGTT AAGCGTCAAG TCTGCATTCT
551  AA

```

The PSORT algorithm predicts inner membrane (0.043).

20 The protein was expressed in *E.coli* and purified as a his-tagged-fusion product (Figure 104A) and also as a GST-fusion (Figure 104B). The recombinant proteins were used to immunise mice, whose sera were used in a Western blot and for FACS analysis (Figure 104C; his-tagged).

These experiments show that cp7387 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 105

- 25 The following *C.pneumoniae* protein (PID 4376281) was expressed <SEQ ID 209; cp6281>:

```

1  MFLQFFHPIV FSDQSLSFLP YLGKSSGIIE KCSNIVEHYL HLGGDTSVII
51  TGVSGATFLS VDHALPIKS EKIIKILSYI LILPLILALF IKIVLRIILF
101  FKYRGLILDV KKEDLKKTLT PDQENLSLPL PSPTTLKKIH ALHILVRSGK
151  TYNELIQEGF SFTKITDLGQ APSPKQDIFG SYNSLLPNFY FHSLVSPNI
30  201  SGEERALNYH KEQQEEMAVK LKTMQACSFV FRSLHLPsmQ TKDKKAGFGL
251  LTFFFpwKIYP L*

```

The cp6281 nucleotide sequence <SEQ ID 210> is:

```

1  ATGTTTCTTC AGTTTTTCA TCCTATAGTC TTCTCGGATC AGTCCTTATC
51  TTTTCTTCCT TACCTAGGAA AAAGCTCTGG CATTATTGAA AAATGTTCCA
101  ATATCGTTGA ACACATTTA CATTTGGGAG GAGACACTTC TGTTATCATC
151  ACAGGAGTTT CTGGAGCTAC CTTTCTATCT GTTGATCATG CCCTCCCAAT
201  CTCGAAATCT GAAAAAAATAA TAAAAAATTCT CTCCTATATT TTAATTCTTC
251  CTCTGATTCT AGCTCTCTT ATTAAAGATCG TTTTACGCAT TATCTTATT
301  TTCAAGTATC GTGGTCTAAT CCTAGATGTT AAGAAGGAGG ATTTGAAAAA
351  AACACTTACA CCTGACCAAG AAAACCTCAG TCTTCCTTA CCATCTCCTA
401  CAACATTAAA GAAAATTCT GCGCTACACA TTTTAGTGC G TTCTGGAAAA
451  ACCTATAACG AGCTTATACA AGAAGGGTT TCTTCACTA AAATCACAGA
501  TCTTGGTCAA GCTCCTTCAC CAAAGCAAGA TATTGGCTTC TCTTATAATT
45  551  CCCCTCTCCC TAACCTCTAT TTTCATTCT TGGTATCTGT TCCAAATATT
601  TCAGGCGAGG AACGGGCTCT TAATTATCAT AAAGAACAC AAGAGGAAAT
651  GGCTGTTAAA TTAAAAACAA TGCAAGCGTG TTCTTTGTC TTCCGATCCC
701  TGCAATTACCA TTCAATGCAA ACGAAGGACA AAAAGGCTGG ATTTGGACTA
751  CTGACGTTT TCCCTTGGAA AATCTACCCCC CTATAA

```

The PSORT algorithm predicts inner membrane (0.5373).

- 50 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 105A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 105B) and for FACS analysis.

5 1 MMNYEDAKLR GQAVAILYQI GAIKFGKHIL ASGEETPLVV DMRLVISSPE
 51 VLQTVATLIW RLRPSFNSSL LCGVPYALT LATSISLKYN IPMVLRRKEL
 101 QNVDPDSAIK VEGLFTPQQT CLVINDMVSS GKSIIETAVA LEENGLVVRE
 151 ALVFLDERRKE ACQPLGPQGI KVSSVFTVPT LIKALIAYGK LSSGDLTLAN
 201 KISEILEIES *

The cp6904 nucleotide sequence <SEQ ID 204> is:

10 1 ATGATGAAC TACGAAGATGC AAAATTACGC GGTCAAGCTG TAGCAATTCT
 51 ATACCAAATC GGAGCTATAA AGTCGGAAA ACATATTCTC GCTAGCGGAG
 101 AAGAAAATCC TCTGTATGTA GATATGCGTC TTGTGATCTC CTCTCCAGAA
 151 GTTCTCCAGA CAGTGGCAAC TCTTATTGCG CGCTCCGCC CCTCATTCAA
 201 TAGTAGCTTA CTCTGCGGAG TCCCTTATAC TGCTCTAACCT CTAGCAACCT
 251 CGATCTCTTT AAAATATAAC ATCCCTATGG TATTGCGAAG GAAGGAATTA
 301 CAGAATGTAG ACCCTCGGA CGCTATTAAA GTAGAAGGGT TATTACTCC
 351 AGGACAAAAT TGTTTAGTCA TCAATGATAT GGTTCCCTCA GGAAAATCTA
 401 TAATAGAGAC AGCAGTCGCA CTGGAAGAAA ATGGTCTGGT AGTTCGTGAA
 451 GCATTGGTAT TCTTAGATCG TAGAAAAGAA CGTGTCAAC CACTTGGTCC
 501 ACAGGAAATA AAAGTCAGTT CGGTATTTAC TGTACCCACT CTGATAAAAG
 551 CTTTGATCGC TTATGGGAAG CTAAGCAGTG GTGATCTAAC CCTGGCAAAC
 601 AAAATTCCCG AAAATTCTAGA AATTGAATCT TAA

20 The PSORT algorithm predicts cytoplasm (0.0358).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 102A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 102B) and for FACS analysis.

The cp6904 protein was also identified in the 2D-PAGE experiment.

25 These experiments show that cp6904 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 103

The following *C.pneumoniae* protein (PID 4376964) was expressed <SEQ ID 205; cp6964>:

30 1 MKKLIALIGI FLVPIKGNTN KEHDHAHATVL KAARAKYNLF FVQDVFPVHE
 51 VIEPISPDCV VHYEGWV*

The cp6964 nucleotide sequence <SEQ ID 206> is:

35 1 ATGAAAAAAAT TGATTGCTTT GATAGGGATA TTTCTTGTTC CAATAAAAGG
 51 AAATACCAAT AAGGAACACG ACGCTCACGC GACTGTTTTA AAAGCGGCCA
 101 GAGCAAAGTA TAATTGTTTC TTTGTTCAAGG ATGTTTTCCC TGTACACGAA
 151 GTTATCGAGC CTATTTCTCC CGATTGCCTG GTACATTATG AAGGGTGGGT
 201 TTGA

The PSORT algorithm predicts inner membrane (0.091).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 103A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a 40 Western blot (Figure 103B) and for FACS analysis (Figure 103C).

These experiments show that cp6964 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 104

The following *C.pneumoniae* protein (PID 4377387) was expressed <SEQ ID 207; cp7387>:

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 108A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 108B) and for FACS analysis.

These experiments show that cp7400 is a surface-exposed and immunoaccessible protein, and that it
5 is a useful immunogen. These properties are not evident from the sequence alone.

Example 109

The following *C.pneumoniae* protein (PID 4376395) was expressed <SEQ ID 217; cp6395>:

```

10      1 MENAMSSSFV YNGPSWILKT SVAQEVFKKH GKGIQVLLST SVMLFIGLGV
      51 CAFIFPQYLI VFVLTIALLM LAISLVLFL L IRSVRSSMVD RLWCSEKGYA
      101 LHQHENGPFL DVKRVQQILL RSPYIKVRAL WPSGDIPEDP SQAAVLLLSP
      151 WTPFSSSVDVE ALLPSPQEKE GKYIDPVLPK LSRIERVSSL VFLSAFTLDD
      201 LNEQGVNPLM NNEEFLFFIN KKAREHGIQD LKHEIMSSLE KTGVPLDPSM
      251 SFQVSQAMFS VYRYLQRQL TTSELRCFHL LSCFKGDVVH CLASFENPKD
      301 LADSDFLEAC KNVEWGEFIS ACEKALLKNP QGISIKDLKQ FLVR*

```

15 The cp6395 nucleotide sequence <SEQ ID 218> is:

```

20      1 ATGGAGAACATC CTATGTCATC ATCGTTTGTG TATAATGGGC CTTCGTGGAT
      51 TTTAAAAACG TCAGTAGCTC AGGAGGTATT TAAAAAGCAC GGTAAAGGGGA
      101 TTCAAGGTTCT CTTAAAGTACT TCAGTGATGC TTTTTATAGG TCTTGGAGTC
      151 TGTGCCTTTA TATTTCCTCA ATATCTGATT GTTTTGTGTT TGACTATAGC
      201 TTTGCTTATG CTCGCTATAA GCTTGGTATT GTTTCTCTTA ATACGTTCTG
      251 TACGCTCTC AATGGTAGAT CGTTTGTGGT GTTCTGAAAA AGGATATGCT
      301 CTTCATCAAC ATGAGAACCGG GCCCTTTTG GATGTAAGC GTGTACAGCA
      351 AATTCTTCTA AGATCACCCCT ATATCAAAGT CGGGCTTTA TGGCCGTCTG
      401 GAGATATCCC TGAGGATCCT TCACAAGCTG CGGTTCTATT ACTTTCTCCT
      451 TGGACTTTCT TTTCATCCGT GGATGTAGAG GCTTTATTAC CGAGTCCTCA
      501 AGAAAAGGAG GGTAAAGTATA TAGATCCTGT GCTGCCCTAAG TTGCTCTAGGA
      551 TAGAGAGAGT CTCACTTTA GTGTTTTGAG GTGCATTAC TTTGGATGAC
      601 TAAACGAAC AGGGAGTC A TCCTTGATG AATAATGAGG AATTTTATT
      651 TTTTATAAAAT AAGAAAGC GTGAGCATGG GATTCAAGGAT TAAACACAG
      701 AGATTATGTC TTCAAGTAAAG AAAACAGGAG TGCCATTAGA CCCCTCAATG
      751 AGTTTCAAG TTTCAAAAGC GATGTTTTCT GTATATCGCT ACTTGAGACA
      801 AAGGGATTTA ACGACTTCAG AATTAAGATG TTTTCACCTC TTAAGTTGTT
      851 TAAAGGGGA TGTGGTTCAT TGTGGTTCTT CATTGAAAA CCCTAAAGAT
      901 TTAGCAGATT CTGACTTTT AGAAGCTTGT AAGAACGTGG AATGGGGTGA
      951 GTTTATTCG GCATGTGAGA AGGCTTTT AAAGAATCCG CAAGGAATT
      1001 CCATTAAGGA TCTAAACAA TTTTGTGA GGTAA

```

The PSORT algorithm predicts inner membrane (0.6307).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 109A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure
40 109B) and for FACS analysis.

These experiments show that cp6395 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 110

The following *C.pneumoniae* protein (PID 4376396) was expressed <SEQ ID 219; cp6396>:

```

45      1 MIEFAFPVHT SVTADRIEDR MACRMNKLST LAITSLCVLI SSVCIMIGL
      51 CISGTVGTYA FVVGIIIFSVL ALVACVFFLY FFYFSSEEFK CASSQEFRFL
      101 PIPAVVPSALR SYEYISQDAI NDVIKDTMQL STLSSLLDPE AFFLEFPYFN
      151 SLIVNHSMKE ADRLSREAFL ILLGEITWKD CETKILPWLK DPNIITPDDFW
      201 KLLKDHFDLK DFKKRIATWI RKAYPEIRLP KKHCILDKSIY KGCKKFLLS

```

These experiments show that cp6281 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

**Example 106 and
Example 107**

- 5 The following *C.pneumoniae* protein (PID 4376306) was expressed <SEQ ID 211; cp6306>:

```
1 MGNHETYIHP GVLPSHQAQD VSRSTVYPSR SFIMRRMLMG WNFNRVPSKS
51 SEQLMDGHRI PLIFFGKHHP TISILNVNRF SWLSIFYNGE RGF*
```

The cp6306 nucleotide sequence <SEQ ID 212> is:

```
10 1 ATGGGAAACC ATGAGACCTA TATACATCCA GGAGTGCTCC CGAGTAGTCA
    51 TGCTCAGGAT GTTAGCAGAT CTACAGTTTA CCCCAGTCGA AGTTTTATCA
    101 TGAGACGTAT GCTCATGGGC TGGAAATTCA ATCGTGTTC CTCGAAGAGC
    151 TCCGAGCAGT TAATGGATGG TCATCGCATA CCTCTTATAT TTTTGGGAA
    201 GCATCATCCT ACTATATCTA TTTAAATGT CAATAGATT TCTTGGCTCT
    251 CCATTTTTTA CAATGGAGAA AGGGGGTTTT GA
```

- 15 The PSORT algorithm predicts cytoplasm (0.167).

The following *C.pneumoniae* protein (PID 4376434) was also expressed <SEQ ID 213; cp6434>:

```
1 MSESINRSIH LEASTPFFIK LTNLCESRLV KITSLVISLL ALVGAGVTLV
51 VLFVAGILPL LPVLILEIIL ITVLVLLFCL VLEPYLIEKP SKIKELPKVD
101 ELSVVETDST L*
```

- 20 The cp6434 nucleotide sequence <SEQ ID 214> is:

```
25 1 ATGTCTGAAA GTATTAAACAG AAGCATTCA TTAGAACGCT CTACACCATT
    51 TTTTATAAAAA TTAACGAATC TCTGTGAAAG TAGATTAGTT AAGATCACTT
    101 CTCTTGTTAT TTCTCTATTAA GCTTTAGTGG GTGCGGGAGT CACTCTGTG
    151 GTTTTATTTG TAGCTGGGAT CCTTCCTTTA CTTCCCTGTAC TCATCTTACA
    201 AATTATTTTA ATAACCGTCC TTGTCTTGCT TTTTTGTTTG GTATTGGAAC
    251 CTTATTTAAT AGAAAAAACCT AGTAAAATAA AGGAACCTACC TAAAGTAGAC
    301 GAGCTATCTG TAGTAGAAC GGACAGTACT CTTTAA
```

The PSORT algorithm predicts inner membrane (0.6859).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 106A; 6306 = lanes 2-4; 6434 = lanes 8-10). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 106B & 107) and for FACS analysis.

These experiments show that cp6306 & cp6434 are surface-exposed and immunoaccessible proteins, and that they are useful immunogens. These properties are not evident from the sequences alone.

Example 108

- 35 The following *C.pneumoniae* protein (PID 4377400) was expressed <SEQ ID 215; cp7400>:

```
1 MRVMRFFCLF FLGFLGSFHC VAEDKGVDLF GVWDDNQITE CDDSYMTEGR
51 EEVEKVVDA
```

The cp7400 nucleotide sequence <SEQ ID 216> is:

```
40 1 GTGAGAGTTA TGAGATTTTT TTGTCTATTT TTTCTTGCGGT TCCTAGGATC
    51 TTTTCATTGT GTTGCTGAAG ACAAGGGCGT GGATTTATTT GGAGTCTGGG
    101 ACGATAACCA AATTACAGAG TGTGACGATA GTTACATGAC AGAGGGTCGT
    151 GAAGAGGTTG AAAAGGTAGT GGACGCTTAG
```

The PSORT algorithm predicts periplasmic space (0.924).

-140-

```

5   751 AACTACCACT CAAAATTCTT TGCTAGTGGT AGTTATGACT TTATTGCAA
    801 GCCCCTATTG GAACAAACAA ATGTAGACGG CTACTATTG GAGTTGATC
    851 ATGAGCGTTC TGGAGACTTC TCTCCTCTCA CCTTCATTTC TGGAGAAAAA
    901 ACTGTCCTGCT TAGGTCTTGT TACCAAGCAA ACCCCTACAC TTGAAAATAA
    951 GGATGAGGTC ATTGCTCGCA TACATCAAGC AGCAGACTAC CTGCCCTTGG
   1001 AAAGACTCTC TCTAAGTCCA CAGTGTGGTT TTGCTTCATG TGAAATAGGA
   1051 AATAAAATTAA CAGAAGAAGA GCAATGGGCT AAAGTTGCTC TAGTAAAAGA
   1101 AATTTCGAA GAAGTTTGGG AATAA

```

The PSORT algorithm predicts cytoplasm (0.2171).

- 10 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 111A) and also as a his-tagged product. The his-tag protein was used to immunise mice, whose sera were used in a Western blot (Figure 111B) and for FACS analysis.

These experiments show that cp6408 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

15 Example 112

The following *C.pneumoniae* protein (PID 4376430) was expressed <SEQ ID 223; cp6430>:

```

20   1 MKLYSISSDV DTPWIFQLMS KVDSYFLFLLG NRIKVVSIVM QEPNLIIGKV
    51 ENVRISTIVK ILKILSFLIF PLILIALALH YFLHAKYANH LLVSKILER
    101 PQYVPIPGRS GDTASHYKLT TLVPVSQKNL QAMGSNPLEV EAALRTTKPS
    151 FFCVPAKYRQ IIISSHGIRF SLDLEQLADD INLDSVSWPT EYLNSTMDFC
    201 SKADKRVIQVNQLRTGTYI NSVGKRSLLK FMLQHLFIDG ITQENPEALP
    251 NNTSGRLTLF PSVRYIYSHF TPQNPTIWPO VFFRQGPLDE DRGGGFIELE
    301 QLQELGVRF P ICPSQGPDPNP NFQGFQGIRI YWEDSYQPNA EV*

```

The cp6430 nucleotide sequence <SEQ ID 224> is:

```

25   1 ATGAAACCTTT ATAGCATCTC TTTCAGATGTA GATACACCTT GGATATTTCA
    51 GCTTATGTCA AAGGTAGATT CTTATCTTTT CTTAGGCGGG AATAGAATCA
    101 AGTTTGTATC TATAGTATG CAAGAACCTA ACTTAATTAT TGGAAAAGTA
    151 GAAAACGTTG GGATCTCCAC AATAGTGAAA ATATTAAAGA TTTTATCCTT
    201 CTTAATCTTC CCTCTGATTT TAATCGCTTT AGCCCTACAC TATTCTTCTAC
    251 ATGCTAAATA TGCTAATCAC TTACTTGTAT CTAAGATTAA AGAAAAGAGCT
    301 CCTCAGTATG TGCCTATTCC TGGTCGTTCA GGAGACACGG CGTCTCATTA
    351 TAAATTAACA ACATTGGTTC CAGTATCCCA AAAAATCTA CAAGCTATGG
    401 GATCAAATCC TCTAGAAGTT GAAGCGGCTC TTCGAACCTAC AAAACCTCT
    451 TTTTCTGTG TACCTGCAAAT ACCCGTCAG ATTATAATT CAAGTCACGG
    501 CATTGCTTT TCTTTAGATC TTGAACAACT TGCTGATGAC ATTAATTTAG
    551 ATTGGGTTTC CTGGCCTACG GAGTATCTTA ACTCTACTAT GGATTTTGCG
    601 AGCAAGGCAG ATAAACCGTGT TATACAGAAAT GTACAAATC TGCGGACAGG
    651 AACTTACATA AATTCTGTAG GAAAGCGTAG CCTTTTAAAAA TTTCATGTTAC
    701 AGCACCTATT TATTGATGGG ATCACACAAAG AAAACCTGA AGCCCTTCCT
    751 AACAAATACAT CTGGAAAGACT GACTCTATT CCTAGTGTTC GTTATATCTA
    801 TTCTCATTTT ACTCCACAAA ATCCCTACAAAT ATGGCCGCAA GTCTTTTCA
    851 GACAAGGTCC TCTAGATGAA GATCGAGGAG GAGGATTGAG GATCTTAGAG
    901 CAATTACAAG AGTTAGGAGT TAGGTTTCCA ATTGCCCCCT CTCAAGGACC
    951 AGACAATCCT AATTTCAGG GTTTCAAGG GATTCTGTAC TATTGGGAAG
   1001 ATTCCCTATCA ACCCAATAAG GAGGTTAA

```

The PSORT algorithm predicts inner membrane (0.5140).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 112A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 112B) and for FACS analysis.

- 50 These experiments show that cp6430 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

-139-

251 ENDVQYQRLL HKVCYFSGEF PAMVLGLGSE VPMVLGLPKV PKDLTWEMFM
 301 ENMPVLLQSK REGHWKISLE DVASL*

The cp6396 nucleotide sequence <SEQ ID 220> is:

```

 5   1 ATGATCGAGT TTGCTTTGT TCCTCATACC TCCGTGACAG CGGATCGGAT
 51  TGAGGATCGC ATGGCCTGTC GCATGAACAA GTTGTCTACT TTAGCAATT
100 101 CAAGTCTTG TGTATTGATC AGTCAGTTT GTATTATGAT TGGGATTTA
150 151 TGCATTTCTG GAACGGTTGG GACCTATGCA TTTGTTGAG GAATTATTT
200 201 TTCTGTGCTT GCTTGGTAG CATGTGTTTT CTTTCTTAT TTCTTTATT
250 251 TTTCTTCTGA GGAATTAAAG TGTGCTTCTT CGCAGGAGTT TCGTTTTTG
300 301 CCTATACCAAG CTGTTGGTTTC TGCATTGCGT TCCTATGAAT ACATTTCTCA
350 351 GGACGCTATC AATGACGTTA TAAAAGATAC GATGCAGTTG TCTACCCCTT
400 401 CTTCTCTTT AGATCCCGAA GCTTTTTCT TAGAATTTC TTTATTTAAC
450 451 TCTTGATAG TGAATCATTC GATGAAGGAA GCGGATCGTT TGTCTCGAGA
500 501 GGCTTTTTTG ATTTTATTAG GTGAGATTAC TTGGAAGGAT TGTGAAACAA
550 551 AAATTTGCC ATGGTTGAAA GATCCTAATA TCACTCCTGA TGATTTCTGG
600 601 AAGCTATTAA AAGACCATT CGATTTAAAG GACTTAAGA AGAGGATCGC
650 651 CACTTGGATA CGGAAGGCCT ATCCAGAAAT TAGATTACCG AAGAACGATT
700 701 GTT TAGATAA GTCTATCTAT AAGGGGTGTT GTAAGTTTT ATTACTTCT
750 751 GAGAATGATG TGCAATATCA GAGGTTATTAA CATAAGGTCT GTTATTTCTC
800 801 TGGGGAGTTT CCTGCCATGG TTTAGGTTT GGGAAAGTAA GTGCCTATGG
850 851 TGT TAGGACT CCCTAAGGTT CCCAAGGATC TTACCTGGGA GATGTTTATG
900 901 GAAAATATGC CTGTTCTTCT GCAAAGCAAA AGAGAGGGC ATTGGAAAAT
950 951 CTCCCTGGAA GACGTAGCCT CTCTTAA

```

The PSORT algorithm predicts inner membrane (0.6095).

- 25 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 110A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 110B) and for FACS analysis.

These experiments show that cp6396 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

30 Example 111

The following *C.pneumoniae* protein (PID 4376408) was expressed <SEQ ID 221; cp6408>:

```

 1 MNTSLKRPLK SHFDVVGSFL RPEHLKKTRE SLKEGSISLD QLMQIEDIAI
 51 QDLIKKQKAA GLSFITDGEF RRATWHYDFM WGFHGVGHHR ATEGVFFDGE
35 101 RAMIDDTYL DKISVSHHPF VDHFKFVKAL EDEFITAKQT LPAPAQFLKQ
151 MIFPNNIEVT RKFYPTNQEL IEDIVAGYRK VIRDLYDAGC RYLQLDDCTR
201 GGLVDPRVCS WYGIDEKGLQ DLIQQQYLLIN NLLVIADRPPD LVVNLHVCRG
251 NYHSKFFASG SYDFIAKPLF EQTNVDGYYL EFDHERSGDF SPLTFISGEK
301 TVCLGLVTSK TPTLENKDEV IARIHQAADY LPLERLSLSP QCGFASCEIG
351 NKLTEEEQWA KVALVKEISE EVWK*

```

- 40 The cp6408 nucleotide sequence <SEQ ID 222> is:

```

 1 ATGAATACTT CACTAAAAAG ACCTCTGAAA TCTCATTTG ATGTTGTCGG
 51 TAGTTTTTTG CGTCCTGAGC ATTTAAAAAA AACTAGAGAA AGCCTAAAG
100 101 AAGGCTCTAT TTCTCTAGAT CAACTCATGC AAATTGAGGA TATCGCTATC
150 151 CAAGATTGAA TCAAAAAACA AAAAGCAGCA GGTCTTCTT TTATTACTGA
200 201 TGGAGAATTG CGCAGAGCTA CGTGGCATTAA CGACTTCATG TGGGGTTTC
250 251 ATGGCGTAGG TCACCACAGA GCTACAGAAG GAGTTTTCTT TGATGGAGAA
300 301 CGCGCTATGA TCGATGATAC CTATCTGACA GACAAGATCT CTGTATCTCA
350 351 CCACCCATTT GTGGATCACT TAAATTGTT AAAAGCTCTA GAAGATGAAT
400 401 TTACGACTGC AAAGCAAAT CTTCTGCAC CGGCACAGTT TTTAAAGCAG
450 451 ATGATCTTCC CTAATAATAT AGAGGTCACA CGTAAATTCT ATCCTACAA
500 501 TCAGGAGCTA ATTGAAGATA TTGTTGCAGG TTATCGTAA GTCATCGCG
550 551 ATCTTTATGA TGCTGGCTGC CGCTATCTCC AATTAGATGA CTGTACTCGG
600 601 GGAGGTTTAG TAGACCCTCG AGTCTGTTCG TGGTATGGTA TCGATGAAAA
650 651 AGGTCTTCAA GATCTGATTC AACAAATATCT TCTGATTAAT AATCTTGTAA
700 701 TTGCGAGATCG TCCCCGATGAT CTAGTCGTTA ATTTACATGT ATGCCGTGGG

```

201 CAAGGCTCCA CATTAGATC CTGAAATCTA TAAACTTGGC ATTCCAATTC
 251 TAGCTATTTG CTATGGCATG CAGCTTATGG CTAGAGATTT TGGAGGGACT
 301 GTAAGCCCTG GTGTAGGAGA ATTGGAATAT ACGCCCATCC ATCTGTATCC
 351 TTGTGAGCTC TTCAACACACA TCGTCGACTG CGAATCTCTA GACACAGAGA
 401 TTCGGATGAG CCATCGGGAT CATGTTACGA CAATTCCCTGA AGGATTAAAT
 451 GTAATCGCAT CCACCTCACCA ATGCCGATC TCAGGAATAG AAAATACCAA
 501 ACAACGGTTG TACGGGCTGC AATTTCATCC CGAGGTTCT GACTCCACTC
 551 CAACGGAAA TAAGATTCTA GAAACTTTTG TTCAAGAGAT CTGTTCTGCT
 601 CCCACACTAT GGAATCCCTT GTATATTCAAG CAAGACCTTG TAAGTAAAAT
 651 TCAAGATACC GTTATTGAAG TATTGATGA AGTCGCTCAG TCATTAGACG
 701 TACAATGGTT AGCTCAAGGA ACCATCTACT CAGATGTTAT TGAGTCCCTCA
 751 CGCTCTGGAC ATGCCCTCGA AGTAATAAAA TCACATCATA ATGTAGGGGG
 801 GCTTCCAAAA AATCTTAAGC TGAAGTTAGT CGAGCCCTTA CGTTATTAT
 851 TAAAGATGA AGTCGAATT TTAGGAGAAG CCCTAGGACT TTCTAGCTAT
 901 CTCTGGACCA GGCATCCCTT TCCCTGGACCT GGCTTGACAA TTCTGTGAT
 951 TGGAGAGATC CTTCTGAAAT ATCTAGCCAT TTTACGACGG GCGGACCTCA
 1001 TCTTTATAGA AGAGCTTAGG AAAGCAAAAC TCTACGATAA AATAAGCCAA
 1051 GCCTTGCCTC TATTCTTCC TATAAAATCA GTATCTGAA AAGGAGATTG
 1101 TAGAAGCTAT GGTTATACCA TAGCATTACG TGCTGTAGAA TCTACAGATT
 1151 TCATGACAGG ACGATGGCC TACCTTCCAT GCGATGTTCT CAGTTCTGC
 1201 TCATCGCGAA TTATTAATGA AATACCCGAG GTAAGCCGAG TGGTCTATGA
 1251 TATTCTGAC AAGCCACCAAG CAACTATAGA ATGGGAATAG

The PSORT algorithm predicts cytoplasm (0.0481).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 114A) and also as
 25 a his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used
 in a Western blot (Figure 114B) and for FACS analysis.

These experiments show that cp6440 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 115

30 The following *C.pneumoniae* protein (PID 4376475) was expressed <SEQ ID 229; cp6475>:

1 MNTYTFSPTL QKSFSLFLLE KLD SYFFF GG TRTQILVITP TNIRLAAKKR
 51 GCKVSTIEKI IKISLFSILLP LVIIA FILRY FLHKKFDKQF LCIPKVISNE
 101 DEALLGSRPQ AVEKAVREIS PAFFSIPRKY QLIRIDTPKD DAPSILFPIG
 151 IEEILKDLCI DTLKQSNLFL KREMDFLGHP EEKALFDSC SIEKDQEWSMS
 201 LESKLLITH FLKYL FVSGI EQLNPGFNPE NGRGYFSEIS TAKIHFHQHG
 251 RYGPPIRSSGP IMKEI*

The cp6475 nucleotide sequence <SEQ ID 230> is:

1 ATGAATACCT ATACCTTCTC TCCTACACTT CAGAAAAGCT TCAGCCTATT
 51 TCTTTAGAA AAATTAGACT CTTACTTTTT CTTGGAGGG ACTCGTACAC
 101 AAATCTTAGT CATCACACCA ACCAATATTA GATTAGCAGC TAAAAAAAGA
 151 GGGTGTAGG TTCTACTAT AGAAAAGATA ATCAAGATCC TCTCTTTAT
 201 CCTGCTGCC C TAGTTATCA TTGCCCTTAT ACCTCGCTAT TTCTTACATA
 251 AGAAATTCTGA TAAACAGTTTC TTGTGTATCC CAAAAGTCAT TTCTAACGAA
 301 GACGAAGCTC TTCTTGATC TAGACCACAA GCAGTTGAAA AAGCAGTTCG
 351 AGAAAATATCT CCAGCCTCTC TCTCTATACC AAGAAAATAC CAACTTATTA
 401 GAATCGACAC TCCTAAAGAT GACGCTCCCT CAATCCTTT CCCTATAGGC
 451 ATAGAGATCA TTCTCAAAGA TTTATGTATT GATACACTCA AGCAATCTAA
 501 TCTTTTCTT AAAAGAGAAA TGGATTCTT AGGTCTACCA GAAGAAAAAG
 551 CATTATTCTGA CTCGATATGT TCTATAGAAA AAGATCAAGA ATGGATGAGC
 601 TTGGAAAGTA AAAAACCTTTT AATCACGCA TTCCTAAAGT ATCTCTTTGT
 651 CTCTGGAATC GAACAACTAA ATCCAGGCTT TAACCCAGAG AATGGCCGTG
 701 GGTATTTTTC AGAAATAAGT ACAGCAAAGA TCCATTTCA TCAGCACGGT
 751 CGATATGGGC CAATCCGTTT TTGGGACCC ATCATGAAGG AAATATAAA

The PSORT algorithm predicts inner membrane (0.5373).

Example 113

The following *C.pneumoniae* protein (PID 4376439) was expressed <SEQ ID 225; cp6439>:

5	1 MSYDTLFKNL EKEDSVHKIC NEIFALVPRL NTIACTEAI	KNLPKADIHV
	51 HLPGTITPQL AWILGVKNGF LKWNSWNTN HRLSPKNPH KQYSNIFRNF	
	101 QDICHEKDPM LSVLQYNILN YDFNSFDRVM ATVQGHRFPP GGIQNEEDLL	
	151 LIFNNYLQOC LDDTIVYTEV QQNIRLAHVL YPSLPEKHAR MKFYQILYRA	
	201 SQTFSKHGIT LRFLNCFNKT FAPQINTQEP AQEAQVWLQE VDSTFPGLFV	
	251 GIQSAGSESA PGACPKRAS GYRNAYDSGF GCEAHAGEGI ETRTIFSSAK	
	301 VNPEGLIEIT RVTFSSLRK QPSSLPIRVT CQLG*	

10 The cp6439 nucleotide sequence <SEQ ID 226> is:

15	1 ATGTCTTATG ATACGTTATT CAAGAATCTT GAAAAGGAAG ATTCTGTACA	
	51 TAAGATATGC AATGAGATCT TTGCAATTAGT ACCACGACTC AATACAATCG	
	101 CTTGACCGA AGCTATCATC AAAAACCTCC CCAAAGCAGA TATCCATGTA	
	151 CACCTTCCTG GGACCTAAC ACCTCAATT A GCTGGATT TAGGTGTGAA	
	201 AAATGGGTTT TTAAAATGGT CTATAATTCT TTGGACCAAT CATCGATTAC	
	251 TTCTCCTAA GAATCCTCAT AAACAATACT CCAATATTTT CCGAAACTTT	
	301 CAAGATATCT GTCACGAAAA GGATCCGGAT TAAAGTGTAT TACAATATAA	
	351 TATCTTAAAT TAGGATTTTA ATAGCTTTGA TAGAGTGTAT GCTACAGTAC	
	401 AAGGACATCG CTTTCTCCT GGAGGAATCC AAAATGAAGA AGACCTTCCTT	
	451 CTCATTTCA ATAACATATCT CCAGCAATGT CTGGACGATA CTATCGTGTA	
	501 TACTGAAGTA CAACAAAATA TCCGCCCTTGC CCATGTTTG TATCCTTCAT	
	551 TACCTGAAAA GCACGCCCGT ATGAAGTTTT ATCAAATCTT GTATCGTGCT	
	601 TCGCAAACGT TTTCAAAACA CGGGATTACT TTACGATTTT TAAACTGCTT	
	651 CAATAAAACA TTTGCTCCAC AAATAAACAC ACAAGAACCT GCCCAAGAAG	
	701 CTGTTCAATG GCTCCAAGAG GTTGATTCTA CATTTCCTGG TCTATTTGTA	
	751 GGGATACAAT CCGCAGGATC AGAATCTGCG CCCGGAGCCT GTCCTAAGCG	
	801 ATTAGCTTCT GGATATAGAA ATGCTTATGA CTCAGGGTTT GGTTGTGAAG	
	851 CTCATGCTGG AGAAGGCATA GAGACCGGA CTATTTTTC GTCAGCTAAG	
	901 GTAATCCAG AGGGATTGAT CGAGATAACC CGAGTGACTT TCTCGTCTCT	
	951 TAAACGAAAA CAGCCATCTA GTTTACCCAT AAGAGTTACT TGCCAGTTAG	
	1001 GATAA	

The PSORT algorithm predicts cytoplasm (0.1628).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 113A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 35 113B) and for FACS analysis.

These experiments show that cp6439 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 114

The following *C.pneumoniae* protein (PID 4376440) was expressed <SEQ ID 227; cp6440>:

40	1 LQSARRHLNT IFILDGFSQL TYVLAKQVRK LFVYCEVLPW NISVQCLKER	
	51 APLGIILSGG PHSVYENKAP HLDPEIYKLG IPIILAICYGM QLMARDFGGT	
	101 VSPGVGEFGY TPIHLYPCEL FKHIVDCESL DTEIRMASHRD HVTTIPEGFN	
	151 VIASTSQCSI SGIENTKQRL YGLQFHPEVS DSTPTGNKIL ETFVQEICSA	
	201 PTLWNPLYIQ QDLVSKIQDT VIEVFDEVAQ SLDVQWLAQG TIYSDVIESS	
	251 RSGHASEVIK SHHNVGGLPK NLKLKLVEPL RYLFKDEVRI LGEALGLSSY	
	301 LLDRHPFPGP GLTIRVIGEI LPEYLAILRR ADLIFIEELR KAKLYDKISQ	
	351 AFALFLPIKS VSVKGDCRSY GYTIALRAVE STDFMTGRWA YLPCDVLSQC	
	401 SSRIINEIPE VSRVVYDISD KPPATIEW*	

The cp6440 nucleotide sequence <SEQ ID 228> is:

50	1 TTGCAGAGTG CAAGGAGACA TTTGAACACC ATATTTATTCT TAGATTTTGG	
	51 ATCTCAATAT ACTTATGTTAGT TAGCAAAGCA AGTGCAGGAAG TTATTTGTAT	
	101 ATTGCGAAGT TCTTCCCTGG AATATCTCTG TGCAATGTTT AAAAGAAAGA	
	151 GCGCCTTGG GGATCATTCT CTCAGGAGGT CCTCACTCTG TCTATGAAAA	

-144-

351 ALLVRKLQFR GAIKSAYFEK LTEIEKELRS LQDVVIKSLEL ELIHKIKDIV
 401 TEET*

The cp6486 nucleotide sequence <SEQ ID 234> is:

```

  1 GTGGTGGTTG TCGCTTATT TATCCTGGG ATTTCTTT TATCTGGTTC
  5 51 TCTGCATTC CTTGTTCAT ACGCTTGCAG AGTTCTTTA GGAGCGGCGC
 10 101 TTCCCATACT TTGCATAGGT CTTGTTTAT TGGCTGTAGC TCTTATTGTT
 15 151 TTCTTATGTC ACAAACACAA GACTCGTAA GATTTAGATT ATTATGATCA
 20 201 AGATTTAGAT TCTTGGTGA TTCATAAGAA AGAGATCCCC AATGACATCT
 25 251 CTGAGTTGCG GGTAACATTT GAAAAGTGC AAAATCTGTT TCAGTCCAT
 30 301 ACGAAAGATT TCTCTGATCT AAGCCAAGAG CTTCAAGGTA AATTATATCAA
 35 351 TTGCATGGAG AAATGGCTAA CTTTAAAGA CGAAGTGACT AAATTCTTA
 40 401 TTGTTGAGA TAGATTTTA GAAACCAGAA GAAATTTCAC CACTTTGGA
 45 451 GAACAGGTTA AAGGGATCCA AAGCAATATT TTTGATTTGC ATGAGGAAAA
 50 501 GTCTTCATTA TATTTAGAAT TGTATAGGCT TAGGAAAGAC CTCCAAGTTC
 55 551 TATTAATTTT TTTTCTGTC CCCCCAGGTA TACTCAAGGT AGATTATGAT
 60 601 GAAATTGAGG CTATCAAAGG TCTGTTTATA AGATTAACCT CTAGATTAGA
 65 651 TAAGCTTGAT GTGAAAGCTC AGGAACGTAA GAAGTTCATT AATGAAATGA
 70 701 GTAGGAAATT TAAAGAAGTA GAGAAAGCTT TTGATATTGT CGATAGGGCA
 75 751 ACAAAAAAGC TTATGGATAG AGCCAAGAAA GAAAGTCCGG CACGTCTTT
 80 801 CATGGGTAGA ACTGAGTCTC TCTTAGAAAT GAAAAAAAT GAAGAAGCCC
 85 851 TTAAAAATCA GGGGCTAGAT CCTGAAAATC TTTCCCATCC TGAACCTTT
 90 901 AGTCCGTATC AACAGCTTT AATTGAAAT TATTTAAATA GCGAAATAGT
 95 951 TCTGCATCAT TATGAGTTCC TTATTTCTGG AACAGTAAC TCTGGCCTAA
 100 1001 CTCTTGAAGA ATGTGAAAAT CGAATGAGGG CGGGCTCTAC TGGGTTGAAAC
 105 1051 GCCCTTCTGG TCGTAAAGCT CCAGTCAGA GGTGCTATAA AATCTGCGTA
 110 1101 TTTTGAAAAA CTCACAGAGA TTGAAAAGA GTTACGATCA CTTCAAGACG
 115 1151 TAATAAAGTC ATTGGAACTA GAACTGATCC ATAAGATAAA AGATATAGTG
 120 1201 ACAGAAGAAA CTTAG

```

The PSORT algorithm predicts inner membrane (0.7474).

30 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 117A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 117B) and for FACS analysis.

These experiments show that cp6486 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

35 Example 118

The following *C.pneumoniae* protein (PID 4376526) was expressed <SEQ ID 235; cp6526>:

```

  1 MSPFKKIVNRLLCYISFQKE SRTLPIIIRE PRMTTKSLS FNSVISKNKI
  5 51 HFIISLGCSRNLVDSEVMIGI LLKAGYESTN EIEDADYLIL NTC AFLKSAR
 10 101 DEAKDYLDHL IDVKKENAKI IVTGCMTSNH KDELKPWMSH IHYLLGSGDV
 15 151 ENILSAIESR ESGEKISAKS YIEMGEVPRQ LSTPKHYAYL KVAEGCRKRC
 20 201 AFCIIPSIKG KLRSKPLDQI LKEFRILVNK SVKEILLIAQ DLGDYGKDLS
 25 251 TD RSSQLES LHELLKEPGD YWLRLMLYLYP DEVSDGIIDL MQSNPKLLPY
 30 301 VDIPLQHIND RILKQMRRTT SREQILGFLE KLRAKVPQVY IRSVIVGFP
 35 351 GETQEEOFQEL ADFIGEGWID NLGIFLYSQE ANTPAAELPD QIPEKVKESR
 40 401 LKILSQIQKR NVDKHNQKLI GEKIEAVIDN YHPETNLLLT ARFYGQAPEV
 45 451 DPCIIIVNEAK LVSHFGERCF IEITGTAGYD LVGRVVKKSQ NQALLKTSKA
 50 501 *

```

The cp6526 nucleotide sequence <SEQ ID 236> is:

```

  1 ATGAGTCCTT TTAAGAAAAT AGTAAATCGC TTACTATGCT ATATTCTTT
  5 51 TCAAAAAGAA TCAAGAACTC TCCCAATCAT TATTAGAGAA CCTAGGATGA
 10 101 CAACAAAAAG TTTAGGATCT TTCAATTCAAG TTATTTCCA AAATAAAATT
 15 151 CATTTTATTA GTTGGGATG CTCTCGGAAC CTTGAGATA GCGAAGTCAT
 20 201 GCTAGGCATT CTTCTTAAGG CAGTTACGA GTCTACTAAT GAAATTGAAG
 25 251 ATGCTGACTA TTTAATTAA AATACCTGTG CGTTTTAAA AAGTGTAGA
 30 301 GATGAAGCTA AAGATTATCT AGACCATCTA ATTGATGTAA AAAAGAGAA

```

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 115A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 115B) and for FACS analysis.

These experiments show that cp6475 is a surface-exposed and immunoaccessible protein, and that it
5 is a useful immunogen. These properties are not evident from the sequence alone.

Example 116

The following *C.pneumoniae* protein (PID 4376482) was expressed <SEQ ID 231; cp6482>:

10	1	MLVELEALKR	EFAHLKDQKP	TSDQEITSLY	QCLDHLEFVL	LGLGQDKFLK
	51	ATEDEDVLFE	SQKAIDAWNA	LLTKARDVLG	LDIGAIYQT	IEFLGAYLSK
	101	VNRRAFCIAS	EIHFLKTAIR	DLNAYYLLDF	RWPLCKIEEF	VDWGNDCVEI
	151	AKRKLCIFEK	ETKELNESLL	REEHAMEKCS	IQDLQRKLSD	IIIELHDVSL
	201	FCFSKTPSQE	EYQKDCLYQS	RLRYLLLHYE	YTLLCKTSTD	FQEQRARKEE
	251	FIREKFSLLE	LEKGIKQTKE	LEFAIAKSKL	ERGCLVMRKY	EAAAKHSLDS
	301	MFEEETVKSP	RKDTE*			

15 The cp6482 nucleotide sequence <SEQ ID 232> is:

20	1	ATGCTAGTAG	AGTTAGAGGC	TCTTAAAAGA	GAGTTTGCGC	ATTTAAAAGA
	51	CCAGAACCG	ACAAGTGACC	AAGAGATCAC	TTCACTTTAT	CAATGTTGG
	101	ATCATCTTGA	ATT CGTTTA	CTCGGGCTGG	GCCAGGACAA	ATTTTTAAAG
	151	GCTACGGAAG	ATGAAGATGT	GCTTTTTGAG	TCTCAAAAAG	CAATCGATGC
	201	GTGGAATGCT	TTATTGACAA	AAGCCAGAGA	TGTTTTAGGT	CTTGGGGACA
	251	TAGGTGCTAT	CTATCAGACT	ATAGAATTCT	TGGGTGCCCA	TTTATCAAAA
	301	GTGAATCGGA	GGGCTTTTG	TATTGCTTCG	GAGATACATT	TTCTAAAAAC
	351	AGCAATCGA	GATTGAATG	CATATTACCT	GTTAGATTTT	AGATGGCCTC
	401	TTTGAAGAT	AGAAGAGTTT	GTGGATTGGG	GAATGATG	TGTTGAAATA
25	451	GCAAAAGGAA	AGCTATGCAC	TTTTGAAAAA	GAAACCAAGG	AGCTCAATGA
	501	GAGCCCTCTT	AGAGAGGAGC	ATGCCATGGA	GAAATGCTCG	ATTCAAGATC
	551	TGCAAAGGAA	ACTTAGGCAC	ATTATTATTG	AATTGCATGA	TGTTTCTCTT
	601	TTTGTGTTT	CTAAGACTCC	CAGTCAAGAG	GAGTATCAA	AGGATTGTTT
	651	GTATCAATCA	CGATTGAGGT	ACTTATTGTT	GCTGTATGAG	TATACATTGT
30	701	TATGTAAGAC	ATCCACAGAT	TTCAAGAGC	AGGCTAGGGC	TAAAGAGGAG
	751	TTCATTTAGGG	AGAAAATTCA	CCTTCTAGAG	CTCGAAAAGG	GAATAAAACA
	801	AACTAAAGAG	CTTGAGTTT	CAATTGCTAA	AAGTAAGTTA	GAACGGGGCT
	851	GTTTAGTTAT	GAGGAAGTAT	GAAGCTGCCG	CTAAACATAG	TTTAGATTCT
	901	ATGTTCGAAG	AAGAAACTGT	GAAGTCGCCG	CGGAAAGACA	CAGAATAA

35 The PSORT algorithm predicts cytoplasm (0.4607).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 116A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 116B) and for FACS analysis.

These experiments show that cp6482 is a surface-exposed and immunoaccessible protein, and that it
40 is a useful immunogen. These properties are not evident from the sequence alone.

Example 117

The following *C.pneumoniae* protein (PID 4376486) was expressed <SEQ ID 233; cp6486>:

45	1	VVVVALFILG	IFFLSGSLAF	LVHTSCGVLL	GAALPILCIG	LVLLAVALIV
	51	FLCHKHKTRO	DLDYYDQDLD	SLVIHKKEIP	NDISELRVT	EKLQLNLFQFH
	101	TKDFSDLSQE	LQGKFINCME	KWLTLDEDEV	KFLIVRDRFL	ETRRNFTTFG
	151	EQVKGIQSNI	FDLHEEKSSL	YLELYRLRKD	LQVLLNNFL	PPGILKVVDYD
	201	EIEAIKGLFI	RLTSRLDKLD	VKAQERKKFI	NEMSREFKEV	EKAFDIVDRA
	251	TKKLMDRAKK	ESPARLFMGR	TESLLEMKKN	EEALKNQGLD	PENLSHPELF
	301	SPYQQLLILN	YLNSEIVLHH	YEFLISGTVT	SGLTLEECE	RMRAASTGLN

The PSORT algorithm predicts cytoplasm (0.1668).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 119A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 119B) and for FACS analysis.

- 5 These experiments show that cp6528 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 120

The following *C.pneumoniae* protein (PID 4376627) was expressed <SEQ ID 239; cp6627>:

```

10      1 MKCSPLTLVP HIFLKNDCEC HRSCSLKIRT IARLILGLVL ALVSALSFVF
      51 LAAPISYAIAG GTLALAAIVI LITLTVVALL AKSKVLPPIPNE LQKIIYNR
     101 PKEVFYFVKT HSLTVNELKI FINCWKS GTD LPPNLHKKAE AFGIDILKSI
     151 DLTLFPEFEE ILLQNCPFLYW LSHFIDKTES VAGEIGLNKT QKVYGLLGPL
     201 AFHKGYTTIE HSYTRPILLT ISESQYKFY SKASKNQWDS PSVKKTC EEI
     251 FKELPHNMIF RKDVQGISQF LFLLFFSHGIT WEQAQMQLI NPDNWKMLCQ
     301 FDKAGGHCSM ATFGGFLNTE TNMFDPVSSN YEPPTVNFMTW KELKVLLEKV
     351 KESPMHPASA LVQKICVNNT HHQNLKRWQ FVRNTSSQWT SSLPQYAFHA
     401 QTYKLEKKIE SSLPIRSSL*

```

The cp6627 nucleotide sequence <SEQ ID 240> is:

```

20      1 ATGAAGTGTA GTCCTTAAC ACTAGTTCCC CATATATT TT TAAAAAATGA
      51 CTGCGAATGT CATAGATCTT GTTCTTTAAA ATTAGGACA ATTGCCGAC
     101 TCATTCTTGG GCTTGTTCTA GCTCTTGT TA GCGCACTTTC TTTTGTMTTC
     151 CTTGCTGCGC CGATTAGCTA TGCTATTGGA GGAACATTAG CTTTAGCCGC
     201 TATCGTAATC TTGATTATAA CGCTAGTCGT AGCACTGCTA GCTAAATCAA
     251 AGGTTCTGCC CATCCCCAAC GAACCTCAGA AGATTATT TA CAATCGCTAT
     301 CCTAAAGAACG TCTTTTATTT CGTGAAAACA CACTCCCTGA CTGTTAACGA
     351 ATTAAAAATA TTATTAATT GCTGGAAAAG CGGTACAGAC CTGCCTCCGA
     401 ATTACATAAA AAAAGCAGAG GCTTCGGGA TCGATATTCT AAAATCTATA
     451 GATTTAACCC TGTTTCCAGA GTTCAAGAG ATTCTTCTTC AAAACTGCC
     501 GTTATACTGG CTCTCCCTT TTATAGACAA AACTGAATCT GTTGTGGGG
     551 AAATCGGATT AAATAAAACA CAAAAAGTTT ATGGTTTACT TGGGCCCTTA
     601 GCGTTTCATA AAGGATATAC AACTATTTT CACTCTTATA CACGCCCTCT
     651 ACTAACATTA ATCTCAGAAT CACAGTATAA GTTCCCTATAT AGTAAAGCGT
     701 CTAAGAACATCA ATGGGATTCT CCTTCTGTGA AAAAACCTG CGAAGAAATA
     751 TTCAGGAAC TCCCCCACAA TATGATTTT CGGAAGGATG TTCAAGGAAT
     801 CTCACAATTTC TTATTTCTTT TCTTTCTCA TGGTATCACT TGGGAACAGG
     851 CTCAGATGAT TCAACTTATA AACCTCTGATA ATTGGAAAAT GTTGTGTCA
     901 TTTGATAAAAG CAGGAGGCCA CTGTTCCATG GCAACATTTG GAGGCTTTT
     951 GAATACTGAA ACAAAATATGT TCGATCCAGT ATCCTCTAAC TATGAACCTA
    1001 CAGTGAACCT CATGACGTGG AAAAGAATTGA AGGTTTTACT AGAGAAAAGTA
    1051 AAAGAAAGTC CTATGCACCC AGCGAGTGCT CTTGTTCTAGA AGATATGCGT
    1101 AAATACAACG CACCATCAAA ATCTGTTAAA ACGATGGCAA TTTGTTCTGA
    1151 ATACGAGTTC ACAATGGACA TCAAGCTTAC CTCAGTATGC TTTCCACGCC
    1201 CAAACCTACA AACTAGAGAA AAAATAGAA AGCAGTCTCC CTATACGATC
    1251 TTCCCTATAA

```

- 45 The PSORT algorithm predicts inner membrane (0.7198).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 120A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 120B) and for FACS analysis.

- These experiments show that cp6627 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

5 351 CGCTAAAATT ATTGTAAC TG GATGCATGAC TTCCAACCAC AAAGATGAGC
 401 TTAAACCCCTG GATGTCACAC ATCCATTACC TACTAGGTTC TGGGGATGTT
 451 GAGAATATTG TTTCTGCTAT TGAGTCTCGT GAATCTGGAG AAAAAATCTC
 501 TGCAGAACAGT TACATTGAGA TGGGAGAAGT TCCAAGACAG CTTTCCACAC
 551 CAAAACACTA TGCCTATTGA AAAGTTGCTG AGGGCTGTAG AAAACGTTGT
 601 GCTTTTGTA TTATTCCTTC CATTAAAGGA AAGCTCCGCA GCAAACCTCT
 651 CGATCAAATT CTTAAAGAAT TCCCACATCCT TGTAAACAAG AGTGTGAAAG
 701 AGATTATATT GATAGCTCAA GACCTAGGAG ATTATGGAAA GGATCTCTCT
 751 ACAGACCGCA GTTCGCGACT AGAACACTA TTACATGAGT TACTGAAAGA
 801 GCCTGGGTGAT TATTGGCTGC GGATGTTGTA TTATATCCT GATGAAGTGA
 851 GTGATGGCAT TATAGATCTT ATGCAATCTA ATCCCCAAACT TCTTCCTAT
 901 GTAGATATTG CCTTACAGCA CATTAAACGAC CGTATTTAA AGCAAATGCG
 951 AAGAACGACT TCTAGGGAGC AAATCCTAGG ATTCCCTAGAA AAATTACGTG
 10 1001 CCAAGGTTCC TCAGGTCTAT ATCCGTTCTT CTGTTATTGT GGGTTTCCCC
 1051 GGTGAAACTC AGGAAGAATT CCAGGAGTTA GCTGATTTA TTGGTGAGGG
 1101 TTGGATTGAT AATCTCGGAA TTTTCTTGTA CTCTCAAGAA GCGAATACCC
 1151 CGGCAGCAGA ACTCCCTGAC CAGATACCAG AAAAAGTTAA AGAATCGAGG
 15 1201 TTGAAAATTG TATCTCAAAAT TCAGAACACG AATGTGGATA AACATAATCA
 1251 GAAGCTCATT GGGGAAAAAA TAGAACAGT TATTGATAAC TATCATCCTG
 1301 AAACGAATCT TTTACTCACT GCAAGGTTCT ATGGACAAGC TCCTGAAAGTG
 1351 GACCCCTTGTA TTATTGTAAGA TGAGGCGAAG CTTGTTTCTC ATTTTGGAGA
 1401 AAGATGCTTT ATAGAAATCA CAGGGACTGC TGGTTACGAC CTTGTAGGGC
 1451 GTGTTGTAAGA AAAATCTCAG AACCAAGCTT TGCTAAAAAC TAGCAAAGCT
 1501 TAG

25 The PSORT algorithm predicts cytoplasm (0.1296).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 118A) and also as a his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 118B) and for FACS analysis.

30 These experiments show that cp6526 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 119

The following *C.pneumoniae* protein (PID 4376528) was expressed <SEQ ID 237; cp6528>:

35 1 MKNNINNNEC YFKLDSTVDG DLLAANLKTG DTQAGQGISST ETFSVQGNAT
 51 FKDQVSATGL TSGTTYNLNA QNFTSSQISI DFKNRNLSCN ALPKEDCDPV
 101 PANYVRSPEY FFCSKPLIGD FDFNSGESYL PLTGSEYTLQ QSRNWNSIFR
 151 FIGWKQSTRE LTVGGNTAIQ FLAAGTYIVS FTVGKRWGN NGWGGAIYIN
 201 NGLGQVQCES TIYSGGGYAT IGLTGTISYR ASVDVAPNPN DPNASDRYRA
 251 GIFYLSNGGS SAGIGNYSFS LLYPYPDDRG*

The cp6528 nucleotide sequence <SEQ ID 238> is:

40 1 ATGAAAAACAA ATATTAATAA TAATGAGTGC TATTTTAAAT TAGACTCAAC
 51 TGTAGATGGT GATTGTTAG CAGCCAATCT CAAGACCTTT GATAACACAGG
 101 CCCAAGGAAT CTCATCGACT GAAACATTCTT CTGTTCAAGGG GAATGCAACA
 151 TTAAAGATC AAGTTTCAGC AACTGGATTA ACTTCAGGGAA CTACTTATAAA
 201 TTAAATGCA CAAAACCTTA CTTCCCTCCCA AATCTCTATA GATTTTAAAA
 251 ATAATCGTCT GAGTAATTGT GCATTGCCAA AAGAAGACTG CGATCCGGTG
 301 CCAGCGAATT ATGTCGTT CCCCCAATAT TTTTCTGTT CCAAGCCTCT
 351 GATCGGAGAT TTTGATTGTTA ACTCAGGGGA ATCTTATTG CCTCTGACTG
 401 GTTCGGAATA TACTCTATAT CAGTCACGTG ATGTAATAG TATATTTCGT
 451 TTATAGGAT GGAAGCAAG TACACGAGAA TTAACGTAG GGGGAAATAC
 501 TGGCATAACAA TTTCTGCCAG CAGGAACCTA TATCGTTCA TTTACTGTTG
 551 GTAAACGGTG GGGATGGAAT ATGGTTGGG GAGGAGGCOAT TTATATCAAT
 601 AATGGTTTAG GACAAGTCCA ATGTGAAAGC ACGATTTATA GTGGTGGAGG
 651 GTATGCAACA ATAGGTACAC TGGGGACCTC AATATATAGA GCCTCTGTAG
 701 ATGTAGCTCC TAATCCTAAAT GATCCGAATG CTTCGGATCG CTATAGAGCG
 751 GGTATTTCT ATCTCAGTAA CGGTGGTTCT AGTGCAGGTA TAGGGAATTA
 801 CTCCCTTTCT CTTCTCTATT ATCCGGACGA TAGAGGGTAG

-148-

351 TEEEQWKKIA FVKEIAKEIW G*

The cp6732 nucleotide sequence <SEQ ID 244> is:

	1	ATGGAAATGA	TGAGCCCATT	CCAAACACCT	GAGCAATGTC	ATTTTGATGT
5	51	TGTGGGAAGT	TTCTTACGTC	CTGAAAAGTCT	TACACGAGCA	CGCTCTGATT
	101	TTGAAGAAGG	AAGAATTGTC	TATGAGCAGA	TGCGAGTTGT	CGAAGATGCT
	151	GCTATTGCGTA	ATCTCATAAA	AAAGCAAACA	GAAGCAGGTC	TTATCTTTT
	201	TACTGATGGG	GAATTCCGTA	GGTATAGTTG	GGATTTCGAC	TTTATGTGGG
10	251	GATTCCATGG	CGTGGATCGT	CGCAGGGACT	CTAATGACCC	TGAAATTGGA
	301	GTGTATCTTA	AAGATAAAAAT	CTCCGTATCA	AAACATCCGT	TTATAGAACAA
	351	TTTCGAGTTT	GTCAAAACCT	TTGAGAAGGG	AAATGCAAA	GCAAAACAAA
15	401	CGATTCCCTTC	TCCATCACAA	TTTTTCCATG	AGATGATT	TGCTCTAAT
	451	CTGAAAAATA	CTCGGAAGTT	TTATCCTACG	AATCAAGAGC	TAATTGATGA
	501	TATTGTCTTT	TATTATCGCC	AAGTCATCCA	AGATCTTTAT	GCTGCAGGTT
	551	GTCGTAATTT	GCAGTTGGAC	GATTGTGCTT	GGTGTGCGCT	CTTGGATATA
20	601	CGAGCGCCTT	CTTGGTATGG	TGTTGATTCT	CATGACAGGT	TGCAGGAAAT
	651	TTTAGAACAG	TTTTTATGGA	TCCATAAFTT	AGTGTGAAAG	GATAGACCCG
	701	AGGATCTTTT	TGTAAGTCTG	CATGCTGTC	GTGGTGATTA	TCAGGCCGAG
	751	TTTTCTCTA	GACGAGCTTA	TGATTCTATA	GAGGAGCCCTT	TATTIGCTAA
	801	GACCGATGTC	GATAGTTATC	ACTATTATTG	GGCTCTTGAT	GATAAGTATT
	851	CAGGAGGTC	TGAGCCTTA	GCTTACGTCT	CTGGAGAGAA	ACACGCTCTGC
25	901	TTGGGATGTA	TCTCCAGCAA	CCATTCTTGT	ATTGAAGATC	GAGATGCTGT
	951	GGTTTCTCGT	ATTATGAAG	CTGGAGCTA	CATTCCCTTA	GAGAGACTTT
	1001	CTTGAGGCC	GCAATGTGGG	TTTGCTTCTT	GTGAGGGAGA	CCATAGAATG
	1051	ACTGAAGAAG	AACAGTGGAA	GAAGATCGCC	TTTGTGAAAG	AGATTGCTAA
	1101	AGAGATCTGG	GGATAA			

The PSORT algorithm predicts cytoplasm (0.2196).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 122A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 122B) and for FACS analysis.

30 These experiments show that cp6732 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 123

The following *C.pneumoniae* protein (PID 4376738) was expressed <SEQ ID 245; cp6738>:

35	1	VWLRFLLLVS	YDEKEKDVVV	VCNHSEPNIL	GLPPEAVSQL	IEELSDEGYS
	51	YLNVVRCDSL	GETTVQQRLL	LNADEGRSMT	VVISLPEGH	PDIRNLQLAS
	101	ERIFVSREKE	AADAYASGCK	VVAFDDEHLP	WVSSHIAYAE	EIREKQEQT
	151	QGSLTEEQLG	ALLCNTVSTE	KNLAFALDAV	IKQSVWRFRN	PDLFAYERE
	201	LEASVTDALW	SYVSNLDMP	YTSSQGIVIE	DSSIVRTSQE	HTLIVNCAAF
	251	DKLASQIEFL	CPSDVLPISG	KDPLISDDDE	EELNPKVSSA	ADSKDKT*

40 The cp6738 nucleotide sequence <SEQ ID 246> is:

45	1	GTGTGGCTGC	GCTTTTACT	TTTAGTGTCC	TATGATGAGA	AGGAGAAAGA
	51	CGTAGTTGTC	GTTTGTAATC	ATTCTGAACC	TAATATCCTC	GGCCTGCCTC
	101	CTGAAGCAGT	CTCTCAGCTT	ATTGAAGAGC	TTAGCGATGA	AGGCTATAGC
	151	TATCTGAATG	TAGTGCCTTG	TGAICTCTCC	GGGGAGACTA	CGGTCAACA
	201	ACGTCTGCTA	TTGAATGCCG	ATGAAGGGAG	ATCTATGACG	GTGGTGATCT
	251	CAGAGCTTCC	TGAAGGGCAC	CCCGATATT	GAATTTGCA	GTTGGCATCC
	301	GAAAGAATT	TTGTTTCTCG	TGAAAAAGAA	GCTGCTGATG	CCTATGCTTC
	351	AGGATGTA	GTGGTCGCTT	TCGATGATGA	GCATCTCCCT	TGGGCTCCA
50	401	GTCATATTGC	CTACGCCAG	GAGATCAGAG	AGAAACAAAGA	ACAAACAAATG
	451	CAAGGGTCTT	TAATCTGAAGA	GCAGTTAGGA	GCACTCCTCT	GCAACACAGT
	501	CTCCACAGAG	AAAAATCTAG	CCTTTGCTCT	AGACGCCGTG	ATAAAACAGT
	551	CTGTGTGGAG	ATTCCCGAAT	CCGGATCTTT	TTGCTTATGA	GAGAGAAGCT
	601	CTAGAGGCTT	CAGAACAGA	TGCTTTAGTA	TCTTACGTTT	CAAATTAGA
	651	CATGATACCG	TACACAAGTT	CTCAGGGCAT	AGTCATAGAA	GATAGTAGTA
55	701	TCGTCCGTAC	CTCTCAAGAG	CATACACTCA	TTGTGAACTG	TGCAGCATTC

Example 121

The following *C.pneumoniae* protein (PID 4376629) was expressed <SEQ ID 241; cp6629>:

```

5      1  MSNITSPVIQ NNRSCNYFE LKNSTTIHV ISAILLCGAL IAFCLVAAPV
      51  SYILSGALLG LGLLIALIGV ILGIKKITPM ISSKEQVFPO ELVNRIRAHY
     101  PKFVSDFVSE AKPNLKDLIS FIDLNLQHNS EVGSSTNYNV SEELQQKIDT
     151  FEGIARLKNE VRTASLKRLE SAASSRPLFP SLPKILQKVF PFFWLGEFIS
     201  AGSKVVELHR VKKIGGSLEE DLSDYIKPEM LPTYWLIPLD FRPTNSSILN
     251  LHTLVLARVL TRDVFQHLKY AALNGEWNLN HSDLNTMKQQ LFAKYHAAYQ
     301  SYKHLSQPSL QEDEFYNLL CIFKHRYSWK QMSLIKTVPA DLWENLCCLT
     351  LDHTGRPQDM EFASLIGLTLY TQGLIHKES AFLSSLTLLS LDQFKTIRRQ
     401  STNIAMFLEN LATHNSTFRS LPPITVHPLK RSVFSQPEED ESSILLIG*

```

The cp6629 nucleotide sequence <SEQ ID 242> is:

```

15     1  ATGAGTAATA TAACCTCGCC AGTTATTCAA AATAATCGCT CTTGTAATTA
      51  TTATTTTGAA TTAAAGAATT CAACCACTAT TCATATTGTT ATCAGTGCCA
     101  TCTTACTCTG CGGAGCTTTG ATAGCTTTCT TGTTGTGTAGC AGCTCTGTT
     151  TCCTATATTTC TAAGTGGCGC ATTGTTAGGA TTAGGATTAT TAATAGCCTT
     201  GATTGGTGTG ATTTTAGGAA TAAAAAAAAT CACGCCTATG ATTCATCAA
     251  AAGAACAAAGT ATTCCCCCAA GAACCTGTAA ATAGAACATAG GGCGCACTAT
     301  CCTAAATTTG TCTCTGATTT TGTTTCAGAA GCTAAACCAA ATCTTAAAGA
     351  TCTCATAAGT TTTATTGATC TTCTAAATCA ATTGCACTCT GAAGTTGGAT
     401  CATCTACAAA TTACAACGTA TCTGAAGAAC TACAACAGAA AATAGATACG
     451  TTCGAGGGTA TCGCACGCTT AAAAATGAA GTCCGTACTG CTTCTCTTAA
     501  AAGACTTGAA AGCGCTGCTT CTTCGGTCC CCTCTTCCCC TCTTACCAA
     551  AAATCTTACA AAAGGTATTT CCATTTTCTT GGTAGGAGA GTTTATTTCT
     601  GCAGGCAGCA AGGTTGTAGA GCTCCATCGA GTTAAGAAAA TTGGAGGCAG
     651  CCTCGAAGAA GACCTTAGTG ATTATATAAA ACCAGAGATG CTTCCTACCT
     701  ATGGTTGTAT TCCTTTAGAT TTTAGACCAA CAAATCCCTC TATTCTAAAT
     751  CTACACACAT TAGTTTAGC TAGAGTCTTA ACTCGTGTATG TTTTCAACA
     801  TCTTAAGTAT GCAGCATTAA ATGGCGAGTG GAACCTGAAT CATACTGATC
     851  TAAATACTAT GAAACAGCAG CTCTTGCTA AATATCATGC GGCATATCAA
     901  TCCTATAAAC ATCTATCTCA ACCCTCTCTT CAAGAGGATG AATTCTATAA
     951  CCTGCTCTTG TGTATTTTA AGCATAGGTA CTCGTGGAAG CAGATGTCCT
    1001  TAATAAAAAC AGTCCCAGCT GATTATGGG AAAACCTCTG TTGCTTGACT
    1051  TTAGACCATA CAGGACGACC CCAAGACATG GAATTTGCCT CTCTAATTGG
    1101  TACTCTCTAC ACACAAGGCC TAATTCTAA AGAAAGCGAA GCATTCTTT
    1151  CTTCATTGAC ACTCCTTAGT TTAGATCAGT TTAAACAGAT CCGTCGTCAG
    1201  TCAACCAATA TAGCGATGTT CCTTGAGAAT TTAGCAACTC ATAATCCAC
    1251  CTTTAGAAGC TTACCACCA TAACAGTCCA TCCACTCAAG AGAACCGTCT
    1301  TCTCCCAACC TGAAGAAGAC GAGTCCTCCC TGCTGATAGG TTAG

```

40 The PSORT algorithm predicts inner membrane (0.5776).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 121A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 121B) and for FACS analysis.

45 These experiments show that cp6629 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 122

The following *C.pneumoniae* protein (PID 4376732) was expressed <SEQ ID 243; cp6732>:

```

50     1  MEMMSPFQQP EQCHFDVVGS FLRPESLTRA RSDFEEGRIV YEQMRVVEDA
      51  AIRNLIKKQT EAGLIFFTGD EFRRYSWDFD FMWGFHGVD RRDSNDPEIG
     101  VYLKDKISVS KHPFIEHFEF VKTFEKGNAK AKQTIPSPSQ FFHEMIFAPN
     151  LKNTRKFYPT NQELIDDIVF YYRQVIQDLY AAGCRNLQLD DCAWCRLLDI
     201  RAPSWSYGVDS HDRLQEILOQ FLWIHNLMVK DRPEDLFVSL HVCRGDYQAE
     251  FFSRRRAYDSI EEPLFAKTDV DSYHYYWALD DKYSGGAEPL AYVSGEKHVC
     301  LGЛИSSNHSC IEDRDAVSR IYEAAASYIPL ERLSLSPQCG FASCEGDHRM

```

These experiments show that cp6739 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 125

The following *C.pneumoniae* protein (PID 4376741) was expressed <SEQ ID 249; cp6741>:

5	1 MASCLSAWFS IVREHFYRAF DFSLPFCARI TEFVLGVVIKG IPVGHIIVG
	51 IEWLVSRYLE SFVTKPTFVS DVVSLLKTEK VAGRDHIARV VETLKRQRVA
	101 VAPEDEDKVKH GKIPVHPFGG IQPVEVLTLY PEVQDATLGL AFSKIRNRVR
	151 QAYLQAPRK LQKIYIIGND MNPFEVDDFL HLARLCNETQ RLYPDATISL
	201 YLTASGGRNA MDKKNRKLLS DCELNPKIAC LDFNQGDVVK QATCDCWMVY
	251 HGENDQGTLM QIQEELEKSG EETPWIHVQ KPLSQSLWDF SPFSSLEMKG
10	301 DKEKALEYSE LEKEQLYSRL VYVGERSSVL SLGFGDSRSRG ILMDPKRVHA
	351 PLSEGHHYCHS VLADLENPGQ QKTILAAFLN PKELSSTILQ PISLNLILNS
	401 KTYLRQHFGE FERMRSRDRN VVVVCDSSWW GTDWKEEPSF QHFIMELECR
	451 GYSHFNIFAF RSNSMCVEER RILNESSQEK AFTMIFCEDS VSQGDIRCLH
15	501 LASEGMLCGK ECYAVDVYTS GCANFMMEEV LTLERESNLW NRKHGLWKRE
	551 VRKQKQEAAL DQDESEIYVC NQLTAQQNFA CS*

The cp6741 nucleotide sequence <SEQ ID 250> is:

20	1 ATGGCTTCTT GTTTATCTGC CTGGTTTTCT ATAGTTCTGT AGCACCTTTA
	51 TCGAGCCTTT GATTTTCTT TGCCGTTTG TGCTCGTATT ACGGAATTG
	101 TATTAGGGGT CATCAAGGGG ATCCCTGTTG TGGGTACAT TATTGTTGGG
	151 ATAGAGTGGC TCGTTCTAG GTATTTAGAG AGTTTCGTGA CCAAGCCGAC
	201 ATTGTCCTCT GATGTGGTGA GTCTTCTGAA AACAGAGAAA GTTGTGGTC
	251 GCGATCACAT TGCTCGTGA GTGGAGACTT TGAAGAGGCA GAGAGTCGCT
	301 GTGGCTCCTG AAGATGAGGA TAAGGTCCAT GGGAAAGATTC CTGTGCATCC
25	351 TTTCGGGGGA ATCCAACCTG TAGAAGTTCT CACTCTCAT CCCGAAGTTTC
	401 AAGATGCAAC GTTAGGGCTT GCCTTCTCTA AAATTCGTAA TCGTGTAAAGA
	451 CAGGCGTATT TGCAAGCTCC ACGGCCAAAA CTGAGAGAAA TTTACATCAT
	501 AGGAAACAGAT ATGAATCCTT TTGAAGTTGA CGACTCTTG CATCTAGCCC
	551 GTCTCTGTAA TGAAACTCAA AGACTCTATC CTGAGGCTAC GATTTCCTA
30	601 TATCTAACAG CTTCTGGTGG TCGCAATGCT ATGGACAAAAA AGAATCGGAA
	651 GTTACTTAGT GATTGCGAAC TAAACCCCAA GATTGCTTGT TTGGACTTTA
	701 ATCAGGGTGA TGTAGTCAAA CAAGCAACTT GTGACTGTTG GATGGTGTAT
	751 CATGGGGAGA ATGATCAAGG TACGTTGAAT CAGATTCAAGG AAGAGTTAGA
	801 AAAGTCAGGG GAGGAACCC CTTGGATTCA TGTGGGGCAA AAGCCTCTTT
35	851 CACAATCCTT GTGGGATTTC TCTCCATTTT CATCTTGGT GATGAAGGGAA
	901 GATAAAAGAGA AAGCTCTAGA GTACTCTGAA TTAGAAAAAG AACAGCTATA
	951 TTCTCGATTG GTATACGTAG GAGAGCGCTC TTCGGTTCTT AGTTTGGGGT
	1001 TTGGAGATAG TCGGTCAAGGG ATCTTGATGG ACCCAAAACG GGTGCATGCT
40	1051 CCCTTATCTG AAGGGCATT A TTGTCAATTCC TACCTTGCAG ACTTAGAAAA
	1101 TCCC GGTTA CAAAAAACAA TTTAGCGGC ATTTCTGAAT CCTAAGGAGT
	1151 TGAGCAGTAC CATACTGCAA CCTATATCTC TAAATCTTAT CTTAAATAGC
	1201 AAAACTTACT TAAGGCAGCA CTTGGCTTT TTGAGAGGA TGAGCAGAAG
	1251 TGATCGCAAT GTGGTTGTCG TTGTATGTGA TTCTGGTGG GGTACCGACT
	1301 GGAAGGAGGA GCCAAGCTTC CAACACTTTA TTATGGAGCT AGAGTGTGCA
45	1351 GGGTATTCGC ACTTCATAT TTTGCCCTT AGATCTAATA GCATGTGTGT
	1401 AGAAGAACGT AGGATCTTAA ATGAAAGTTC TCAAGAGAAA GCCTTACCA
	1451 TGATTTCTG TGAGGATTCA GTATCTCAAG GAGATATCCG CTGTTGCAT
	1501 TTGGCGTCTG AAGGAATGCT TTGTGGTAAA GAGTGTATG CTGTCGATGT
	1551 CTATACGTCA GGATGCGCGA ACTTTATGAT GGAAGAAGTC TTAACCTTGG
50	1601 AGCGAGAACATC TAATCTGTGG AATAGAAAGC ATGGTCTTGT GAAAAGAGAA
	1651 GTTAGAAAAAC AGAAACAAAGA AGCTGCTTTG GATCAAGACG AGAGCGAGAT
	1701 TTACGTTGT AATCAGCTGA CGCGCAACA GAACTCGCT TGTCTTGA

The PSORT algorithm predicts inner membrane (0.2869).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 125A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 125B) and for FACS analysis.

751 GATAAGTTAG CGAGCCAAAT AGAGTTCTTA TGCCCCAGTG ACGTGGTGC
 801 CATTTCCTGGT AAAGACCCT TGATTTCTGA TGATGAGGAT GAGGAACGTGA
 851 ATCCTAAAGT TTCATCTGCT GCAGACTCTA AAGATAAAAC CTAG

The PSORT algorithm predicts cytoplasm (0.1587).

- 5 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 123A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 123B) and for FACS analysis.

These experiments show that cp6738 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

10 Example 124

The following *C.pneumoniae* protein (PID 4376739) was expressed <SEQ ID 247; cp6739>:

1 MTHCLHGWF S VVRHHFVQAF NFSRPLYSRI THFALGVIKA IPIVGHLVMG
 51 VDWLISHCFE RGVSHPGFPS DIAPILKVEK JAGRDHISRI ENQLKSLRK
 15 IEVEDLDKVH GQYQENPYAD MASSEVLKLD KGVHVSELGK AFSRVNRIT
 151 RSYSYAPTPQ LDSIAIVGID LVSPEEQENL VRLANEVIQL YPKSKTTLYL
 201 LIDFNKEWVG DISSDKEKQL RSLGLHSEVQ CLSVLEPQGA EGEDTKHFDL
 251 MVGCYGKD SY LREGKILQQA LGTSLGTVPW VNVMHTLPSR YRSRLSLPIN
 301 TEKDKTELYK EISRTHHQLH TLGMGLGAQD SGLLLDRQRL HAPLSQGSCH
 351 HSYLADLTHE ELKILLFSAF VDAKNISKKE LREVSLNFAN DTSVECGCAF
 20 401 YF*

The cp6739 nucleotide sequence <SEQ ID 248> is:

1 ATGACTCAT T GCTTACATGG TTGGTTTTCT GTAGTCGTC ATCACTTTGT
 51 GCAGGCGTTT AATTCTCA C GTCCTTTATA TTCTCGAATT ACCCACTTCG
 101 CTTTAGGGGT GATTAAGGCC ATCCCCATTG TAGGGCATCT TGTTATGGGA
 151 GTCGATTGGT TGATCTCTCA TTGCTTCGAG AGGGGAGTCT CACACCTGG
 201 GTTCCCTTC GATATTGCTC CTATACTGAA AGTAGAAAAG ATCGCGGGCC
 251 GAGATCATAT TTCTAGAAC GAAAATCAGC TAAAGAGCCT TAGGAAAAGT
 301 ATCGAGGTTG AAGATCTAGA TAAAGTCCAC GGGCAATATC AAGAGAATCC
 351 TTATGCAGAT ATGGCCTCTA GTGAGGTTCT TAAACTCGAT AAGGGAGTTC
 401 ATGTTAGCGA GCTTGGCAAA GCCTTTCTA GAGTTCGCAA TCGCATCACC
 451 AGATCCTATA GTTATGCCCT TACTCCTCAG TTGGACTCTA TAGCTATTGT
 501 TGGTATAGAT CTCGTCA GTCAGAACAGA AGAGAATTAA GTACGCTTGG
 551 CGAATGAGGT CATTCAACTC TATCCAAAT CAAAGACAAC TCTATATCTT
 601 CTTATCGATT TTAATAAGGA GTGGGTAGGG GATATCTCCT CTGATAAGGA
 651 AAAACAGCTC CGTTCTCTAG GTCTACATTC TGAAGTCAG TGTCTTCCG
 701 TCTTGGAAAC TCAGGGTGCC GAGGGCGAAG ATACGAAACA CTTTGACCTT
 751 ATGGTCGGCT GTTATGGAA GGATTCTTAC TTAAGGGAGG GTAAAATT
 801 ACAGCAGGCC CTAGGGACTT CGTTAGGTAC TGTTCCCTGG GTGAATGTTA
 851 TGCACACATT GCCATCTAGG TATAGATCTC GGCTTCCCTT ACCTATAAAAT
 901 ACCGAAAAGG ATAAGACAGA GCTTTATAAA GAGATTTCCTC GTACACACCA
 951 TCAGTTGCAT ACTTTGGAA TGGGACTTGG AGCCCAGGAT TCAGGATTGC
 1001 TCTTAGACCG GCAACGACTC CATGCTCCTT TATCTCAAGG GTCTCACTGC
 1051 CATTCCCTATC TTGCGAGATCT CACCCATGAA GAGCTGAAAA TTTTGTATT
 1101 TTCAGCATTG GTGGATGCTA AGAACATAAG TAAGAAAGAG CTTCGTGAGG
 1151 TATCTCTAAA TTTTGCTAAC GATACTTCGG TAGAGTGTGG CTGCGCTTTT
 1201 TACTTTTAG

The PSORT algorithm predicts inner membrane (0.2190).

- 50 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 124A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 124B) and for FACS analysis.

The PSORT algorithm predicts inner membrane (0.2338).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 126A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 126B) and for FACS analysis.

- 5 These experiments show that cp6742 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 127

The following *C.pneumoniae* protein (PID 4376744) was expressed <SEQ ID 253; cp6744>:

10	1 VIQHLLNFAL EETPSISVQY QEQEKLSPCD HSPEIGKKR WNKLESFSTY 51 CSLFMSVKDH YKLNLGIQNS LSGWLLDPYR VCAPLSSPYS CPSYLLDLQN 101 KELRRSLLST FLDPKNLTSE TFRSVSINFG NSSFGQRWSE FLSRVLHDEK 151 EKHVAVVCND AKLLEEGLSP EALSILLEDL RESGYSYLNI LSVSPEGVSK 201 VQERQILRRD LQGRSFTVMI TDPLPLGSEDI RSLQLASDRI LVSSSLDAAD 251 ACASGCKVLV YENPNASWAQ ELENFYKQVE RRR*
----	--

- 15 The cp6744 nucleotide sequence <SEQ ID 254> is:

20	1 GTGATACAAAC ATCTTCTAAA CTTTGCTCTA GAAGAGACCC CTTCCATTTC 51 CGTGCAATAAC CAAGAACAAAG AGAACGCTCTC TCCGTGCGAT CATTCCCCAG 101 AAATAGGTA AAAGAAAAAGA TGGAATAAGC TGGAATCCTT CTCCACGTAT 151 TGGTCTCTGT TTATGTCGTG TAAGGATCAT TATAAGCTGA ATCTAGGAAT 201 TCAGAAATTCC CTGTCAGGGT GGCTTCTGGA TCCCTATAGG GTTTGCAGCGC 251 CTTTATCTTC ACCGTACTCG TGTCCTTCCT ATCTTTAGA TTTGCAAAAC 301 AAAGAGCTAC GTCGTTCCCT TCTGTCAACG TTTCTAGACC CTAAAAATCT 351 CACTAGCGAA ACATTCGTT CTGTCCTCTAT AAACCTTGGC AACTCTTCGT 401 TTGGACAGAG ATGGTCAGAG TTTCTATCTC GTGTTCTGCA CGACGGAGAAA 451 GAAAAGCACG TAGCTGTTGT TTGTAATGAT GCAAAACTTC TGGAAGAAGG 501 ATTGTCCCCA GAGGCATTGT CTCTATTAGA AGAACGACTTA AGAGAACATCAG 551 GGTATTTCGTA TCTAAACATT CTCTCGGTGA GCCCCGAAGG AGTCCTCCAAG 601 GTTCAGGAAC GTCAGATTCT AAGGCGAGAT CTCCAAGGAC GGTCCCTTAC 651 TGTCACTGATT ACAGATCTTC CTMTAGGTAG CGAACGATATC CGTAGTTTAC 701 AATTAGCCTC GGATAGGATT TTAGTCTCCA GTTCTCTTGA TGCCGCCGAT 751 GCATGTGCTT CGGGATGTAAGTCTTAGTC TACGAAAATC CAAATGCATC 801 CTGGGCTCAG GAATTGGAGA ACTTCTACAA ACAAGTTGAG AGAAGAAGG 851 AG
----	---

The PSORT algorithm predicts cytoplasm (0.3833).

- 35 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 127A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 127B) and for FACS analysis.

These experiments show that cp6744 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 128

The following *C.pneumoniae* protein (PID 4376745) was expressed <SEQ ID 255; cp6745>:

45	1 VACPSISSLWF TVVRQHFVNA FDFTHPVCSR ITNFALGIK AIPVLGHIVM 51 GIEWLISWIP RHTVRHGMFT SDVSSAIKVE QTRGHNCLAP LEAYLSSLRV 101 PISQEDLGKV HGRTPEDPFV DITPTEIVQL LPDEELSTVD EALQGVRSRL 151 TYAYRSVEKP MIQDLALVGF GLRDSADLIN FVRLANGVQN HYPHTKVKLY 201 LAKNLADVWD CEISEEEKGQ LRALGLDPKI ESISLTSAGL PSVPEVATVD 251 FMITCYGKDQ EVQDP*
----	--

These experiments show that cp6741 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 126

The following *C.pneumoniae* protein (PID 4376742) was expressed <SEQ ID 251; cp6742>:

5	1 LFVSNFIFFV VMPPIPYISSW ISTVRQHFVK AFDFSRPFC S RVTNFALGVI
	51 KAIPIVGHIV MGMEWLVSSC VAGIIITRSSF TSDVQIVKT EKALGRDHIS
	101 RVAEILQRER GTITPENQDK VHGFVPCPF GRLKSEETLK LKPGEREGLT
	151 DTVFSPIRTR VTRAYLQAPR PEIRTISIVG SKLXTPQDF S QFVSLANETQ
	201 RLHPEALVCL YLTGLNRESQ MCDTTAEEKK QYLHNSGLDS RIQCKDSKED
10	251 DAGSPENPEL WIGYYSREQQ HNIIDQYIQQ CLGKSADPIP WIHVTEDETKD
	301 FYYPPNFTSY SHTRQSTDPT SPPRLPESEG DKDSLYGQLS RSYHHEYMLG
	351 LGLKPEDAGL LMDPDRIYAP LSQGHYCHSY LADIENEDLR TLVLSPFLDP
	401 GNLSSEDLRP VAFNIARLPL ELDLSFFRLV AGQQEGRNIV TLAHGTPRPE
15	451 DLDPDSMNIL TRRLQMSGYS YLNIFSYKSR KMIVKERQFF GDRSEGKSFT
	501 LILFEDPISA ADFRCLQLAA EGMVAKDLP S VADICASGCS CIQFSEMQSP
	551 QAIEYRQWEA RVEDEAGEEA REPVIYSQDQ LSSMLTTQQN FVFSILDAAVK
	601 QAIWRFRSKG LLTMRKA LG EEFILTAIFSY LGSQERNENM GKRTTEEHEV
	651 VISFEELDRM VQVLPAEVPA DSGNDPTRPV PNPDNSPDSS QNEGS*

The cp6742 nucleotide sequence <SEQ ID 252> is:

20	1 TTGTTTGTTT CTAATTTTAT TTTTTTTGTT GTTATGCCAA TTCCCTATAT
	51 TTCTTCTTGG ATTTCCTACCG TTCGACAGCA TTTTGTAAAG GCGTTTGATT
	101 TCTCTCGTCC CTTTTGTTCT AGGGTTACGA ATTTTGCTT AGGGGTCATC
	151 AAGGCCATCC CTATTGTTAG ACATATTGTC ATGGGGATGG AGTGGTTAGT
	201 TTCTTCCTGT GTTGCGGGGA TTATTACTAG GTCTCCCTT ACCTCAGATG
25	251 TCGTTCAAGAT TGTAAAAGACT GAGAACCGT TAGGTGAGA TCATATATCT
	301 CGAGTGGCGG AGATATTGCA AAGAGAAAGG GGGACCATAA CTCCTGAGAA
	351 TCAAGATAAG GTGCATGGGA AGTTTCTGT CTGTCCTTT GGTGTTAA
	401 AATCCGAGGA AACTTTAAAA CTAAAGCCGG GAGAAAGAGA GGGAACTTTA
	451 GATACTGTAT TTTCTCCGAT TCGCACGCG GTGACTCGTG CGTACTTACA
30	501 GGGCCCCCGA CCCGAAATAC GTACGATTTC TATTGTTGGGT TCGAAACTTA
	551 AAACCTCTCA AGATTTCTCG CAATTGTGA GTCTCGCGAA TGAAACGCAG
	601 AGACTGCATC CTGAAGCGTT AGTTGTCTG TATTGACAG GCTTGAATCG
	651 CGAATCTCA G ATGTGCCATA CAACTACTGC AGAGAAGAAG CAGTACCTAC
35	701 ATAACTCAGG TCTCGACTCT AGAAATCCAGT GCAAAGACAG TAAAGAAGAC
	751 GACGCTGGCT CTCCCTGAAAA TCCCGAACCTT TGGATTGGCT ATTATTACAG
	801 AGAGCAACAG CATAATATAG ACGGGCAGTA TATTCAAGCAG TGTCTAGGG
	851 AGAGTGCAGA TCCAATTCCCT TGGATTCATG TTACTGAAGA CACAAAGGAT
	901 TTTTATTACCA CACCAAACCTT TACTTCATAC TCACATACAA GACAATCTAC
40	951 AGACCCAACA TCGCCACCAA GACTCCCTGA AAGTGAGGGG GATAAGGATT
	1001 CCTTGTACGG ACAACTGAGT CGATCGTATC ACCATGAGTA TATGTTGGT
	1051 TTGGGATTAA AACCAAGAGGA TGCAAGACTC CTGATGGACC CGGATAGAAT
	1101 CTATGCTCCT CTATCCCAAG GGCATTATTG TCATTCCCTAC CTTGCGGATA
	1151 TAGAAAATGA GGATCTACGA ACTTTAGTCC TTTCGCCTT CCTAGATCCT
45	1201 GGCATCTTA GTAGCGAGGA TCTTCGTCT GTAGCATTCA ATATCGCTAG
	1251 ATTGCCATTA GAATTGGACT CGTTATTTTT CCGCCTTGT GCGGGTCAGC
	1301 AAGAAGGGAG AAACATAGTT ACCCTTGCCC ACAGGAACCTCC TCGTCCAGAA
	1351 GATCTTGATC CTGACTCAAT GAACATTCTG ACCAGAAGAT TACAAATGTC
	1401 TGGATATAGC TATTTGAACA TTTTCTCTA TAAATCACGG AAAATGATTG
	1451 TAAAAGAACG TCAGTTCTTT GGAGATCGTT CTGAAGGGAA GTCTTTACA
50	1501 TTGATCTTAT TTGAGGATCC CATTAGTGCA GCAGATTTC GTTGTTTGCA
	1551 GCTAGCTGCA GAAGGTATGG TTGCTAAGGA TCTCCCCAGC GTAGCAGATA
	1601 TTGATGCTCCT TGGATCTTCC TGCATTCTAG TTTCTGAGAT GCAGAGTCCT
	1651 CAGGCTATTG AATATAGACA ATGGGAGGCA CGTGTGAAAG ATGAAGCAGG
	1701 AGAAGAACCC AGAGAACAG TAATTATTC TCAGGATCAA TTGAGCAGCA
55	1751 TGCTCACTAC ACAACAGAAT TTGATTTTT CTCTAGATGC TGTGGTAAA
	1801 CAGGCGATCT GGAGATTCCG TTGCAAGAGT CTTCTTACTA TGGAAAGAAA
	1851 GGCACTAGGC GAGGAGTTCT TAACTGCGAT ATTTTCCTAT TTAGGGAGTC
	1901 AGGAGCGTAA TGAGAATATG GGGAAAGAA CTACCGAAGA ACATGAGGTC
	1951 GTTATCAGCT TCGAAGAGCT AGATCGCATG GTGCAAGTCC TCCCAGCCGA
60	2001 AGTCCCTGCA GATTCAAGGCA ATGATCCTAC GCGTCCCGTT CCTAATCCAG
	2051 ATAGTAACCC TGATTCCTCG CAAAATGAAG GCAGTTAG

1101 TGATGAAGAT GTTCCCTCTA CCTCTGAGGA TCCTTCAGAT GATCATCCTT
 1151 CGGATCTTGA AGACTCTTAA

The PSORT algorithm predicts inner membrane (0.1447).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 129A) and also as
 5 a his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used
 in a Western blot (Figure 129B) and for FACS analysis.

These experiments show that cp6747 is a surface-exposed and immunoaccessible protein, and that it
 is a useful immunogen. These properties are not evident from the sequence alone.

Example 130

- 10 The following *C.pneumoniae* protein (PID 4376756) was expressed <SEQ ID 259; cp6756>:

1 MASGIGGGSS LGKIPPKDNG DRSPSPPKG ELGSHEISLP PQEHGEEGAS
 51 GSSHIIHSSSS FLPEDQESQS SSSAASSPGF FSRVRSGVDR ALKSFGNFFS
 101 AEESTSQARET RQAFVRLSKT ITADERRDVD SSSAAATEAR VAEDASVSGE
 151 NPSQGPVPETS SGPEPQRLFS LPSVKKQSGL GRLVQTVRDR IVLPSGAPPT
 201 DSEPLSLYEL NLRLSSLRQE LSDIQSNQDL TPEEKAEATV TIQQLIQITE
 251 FQCGYMEATQ SSVSLAEARF KGVEITSDEIN SLCSELTDPE LQELMSDGDS
 301 LQNLLDETAD DLEAALSHTR LSFSLDDNPT PIDNNPTLIS QEEPIYEEIG
 351 GAADPQR TRE NWSTRLWNQI REALVSSLGM ILSILGSILH RLRIARHAAA
 401 EAVGRCC TCR GEECTSSEED SMSVGSPSEI DETERTGSPH DVPRRN GSPR
 451 EDSPLMNA LV GWAHKHGAKT KESSESSTPE ISISAPIVRG WSQDSSVSFI
 501 VMEDDHIFYD VPRRKDGTYD VPSSPRWSPA RELEEDVFGD YEVPI TS AEP
 551 SKDKNIYMT P RLATPAIYDL PSRP GSSGSS RSPSSDRVRS SSPNRRGVPL
 601 PPVPSPAMSE EGSIYEDMSG ASGAGESDYE DMSRSPSPRG DLDEPIYANT
 651 PEDNPFTQRN IDRILQERSG GASASPVEPI YDEIPWIHGR PPATLPRPEN
 701 TLTNVSLRVS PGFGP EVRAA LLSESVSAVM VEAESIVPPT EPGDGESEYL
 751 EPLGGLVATT KILLQKGWPR GESNA*

The cp6756 nucleotide sequence <SEQ ID 260> is:

1 ATGGCATCAG GAATCGGAGG ATCTAGTGG A TTAGGAAAGA TTCCACCTAA
 51 AGATAATGGG GATAGAAC GTC GATCGCCCTC TCCTAAGGG A GAACTTGGCA
 101 GCCACGAGAT TTCCCTGCCT CCTCAAGAAC ATGGAGAGGA AGGAGCTTCA
 151 GGATCTTCGC ATATACATAG CAGTTCTCT TTTCTACCAAG AAGATCAGGA
 201 GTCTCAGAGC TCTTCTTCGG CAGCTTCTAG CCCGGGATTT TTTTCTCGCG
 251 TACGTTCTGG GGTAGACAGG GCCTTAAAT CATTGGCAA CTTTTTTTCC
 301 GCAGAGTCTA CGAGTCAAGC GCGTGAAACG CGACAAGCTT TTGTTAGATT
 351 ATCAAAAACC ATCACCGCGG ATGAGAGACG GGATGTCGAT TCATCAAGTG
 401 CTGCTGCTAC AGAAGCCCGA GTGGCAGAGG ACGCGAGTGT TTCAGGCAGA
 451 AATCCTTCTC AGGGGGTCTC AGAAACCTCT TCTGGACCA AACCTCAGCG
 501 TTTATTTCTC CTTCTTCAG TAAAAAAACA GAGCGGTTTG GGTGCGGTTGG
 551 TACAGACAGT TCGCGATCGC ATAGTACTTC CTAGTGGGGC TCCACCTACA
 601 GACAGCGAGC CTTTAAGTCT CTACGAGCTA AACCTCCGTT TGAGTAGTTT
 651 ACCTCAGGAG CTCTCTGACA TACAAAGTAA TGATCAGTGTG ACTCCAGAGG
 701 AAAAAGCAGA AGCCACAGTT ACCATACAAAC AGCTGATCCA AATTACAGAA
 751 TTCCAAATGCG GCTATATGG A GGCAACACAA TCTTCGGTAT CTCTAGCAGA
 801 AGCTCGTTT AAGGGGTAG AAACTAGTGA TGAGATCAAT TCCCTCTGTT
 851 CAGAACTGAC AGATCCTGAG CTTCAAGAAC TCATGAGTGA TGGAGACTCT
 901 CTTCAAAACC TATTAGATGA GACTGCCGAC GATTTAGAAG CTGCTTTGTC
 951 CCATACTCGA TTGAGTTTT CTTTAGACGA TAATCCAAT CCGATAGACA
 1001 ATAATCCAAC TCTGATTTCT CAAGAAGAGC CTATTTATGA GGAAATCGGA
 1051 GGAGCTGCAG ATCCCTCAAAG AACCTGGGAA AACTGGTCTA CAAGATTATG
 1101 GAATCAGATT CGCGAGGCTC TGGTTCTCT TTTAGGAATG ATTTTAAGCA
 1151 TTCTAGGGTC CATCTTCGAC AGGTGCGTA TTGCTCGTCA TGCAGCTGCT
 1201 GAAGCAGTGG GTCGTTGTTG CACGTGCCGA GGAGAAGAGT GTACTCTTC
 1251 TGAAGAGGAC TCGATGTCGG TGGGGTCTCC TTCAGAAATT GATGAAACTG
 1301 AAAGAACGGG CTCTCCCAT GACGTTCCAC GCAGAAATGG AAGTCCACGT
 1351 GAAGATTCTC CATTGATGAA TGCCTTAGTA GGATGGGCAC ATAAGCACGG
 1401 TGCTAAAACC AAGGAGAGTT CAGAATCAAG TACCCCCGAA ATTTCGATT
 1451 CTGCTCCCCT AGTGGAGAGGT TGGAGTCAAG ACAGTCCGT CAGTTTATT

The cp6745 nucleotide sequence <SEQ ID 256> is:

```

      1  GTGGCTTGTC CAAGTATTC TTCTTGGTTT ACTGTCGTTG GACAGCATT
      51  TGTAAACGCC TTTGATTTCA CCCATCCCCT TTGTTCTCGG ATTACAAATT
     101  TTGCTTTGGG GATCATTAAG GCAATTCCCG TATTAGGACA CATTGTCATG
     151  GGAATCGAGT GGTTGATTTC CTGGATTCCC AGACACACCG TTGTCATGG
    201  AATGTTTACT TCTGATGTC CTAGTGCTAT TAAAGTAGAA CAAACACGGG
    251  GTCATAATTG TTTAGCTCCC CTAGAAGCCT ATTTAAGTAG CTTGAGAGTC
    301  CCCATTTCCC AAGAAGATCT AGGCAGAAGTA CACGGGAGAA CCCCCAGAAGA
    351  TCCCTTCGTA GATATCACAC CCACAGAAAT TGTCCAACCT CTCCCTGATG
   10  401  AAGAACTCTC TACTGTAGAT GAGGCACTGC AAGGCCTCG TAGTAGGTTA
   451  ACCTATGCCT ATAGGTCCT AGAGAAACCT ATGATTCAAG ATCTTGCTCT
   501  TGTGGGTTTT GGTCTCCGAG ATTCCTGCGGA CCTCTAAAT TTGTCGCGTC
   551  TTGCTAATGG CGTGAGAAT CACTATCCC ATACTAAAGT GAAGCTCTAT
   601  TTAGCGAAGA ACTTGGCAGA TGTCTGGGAC TGTAAAATT CTGAAGAGGA
   651  AAAAGGGCAA CTCCGAGCTC TAGGTTTAAAG CCCTAAAATA GAGAGTATAT
   701  CCCTTACGAG TGCAGGTCTT CCTTCAGTGC CAGAAGTCGC TACTGTCGAT
   751  TTTATGATTA CCTGTTACGG GAAAGATCAG GAAGTCCAAG ATCCCTAG

```

The PSORT algorithm predicts inner membrane (0.2253).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 128A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 128B) and for FACS analysis.

These experiments show that cp6745 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 129

25 The following *C.pneumoniae* protein (PID 4376747) was expressed <SEQ ID 257; cp6747>:

```

      1  MMKQGVGQDA KELYTFLSRG NEHYQPCLWF SLEELGFLF DEKMLCAPLS
      51  EDHYCHSYLV DLVDQHLKDL ILSMFLDPQN ISAGELLKVS INVGDSFSPL
     101  QQKDFLMSVL RDETGNVVV VFKGVLSLPA TQVCKLVEEL NSKDYSYLN
     151  FSCHGDSSPQ LLFRKELEG TSGRYFTVICA LYLGDTDMRS LQLASERIMV
    201  SREFDLVDAY AARCKLLKID HTNWRPGTFS RHADFADAVD VSAGFNSREF
    251  KLITQANQGI LESGELPLPS KTFWEGFLAF CDRVTVTRHF IPMLDAAIKQ
    301  AVWTHKHPSL IDKECEALDL KTQCLPSIVS YLEYVTSNSHE KTSKGPFIQK
    351  EIIADCSPLIK EALFPGSDED VPSTSEDPSD DHPSDLED*

```

The cp6747 nucleotide sequence <SEQ ID 258> is:

```

      1  ATGATGAAAC AAGGAGTCGG GCAGGATGCT AAAGAGCTAT ACACATTTCT
      51  ATCTCGTGGG AATGAGCATT ACCAACCGTG TCTATGGTTC AGTCTCGAAG
     101  AGGAACCTCGG ATTCCCTTTTC GATGAAAAAA TGCTCTCGCG CCCCTCTATCT
     151  GAGGATCACT ATTGCCACTC GTATCTGTA GATCTAGTGG ATCAACATTT
    201  AAAGGATTAA ATATTATCGA TGTGTTTAAAG TCCTCAGAAT ATCTCAGCAG
    251  GAGAACTCCT CAAGGTCTCT ATAAACGTTG GAGATTCTTT TTCTCCTCTA
    301  CAACAGAAAG ATTCCTCTC GATGGTCTTA CGTGATGAAA CGGGAAAAAA
    351  CGTCGTCGTG GTTTTAAAG GAGTTCTCTC CTTACCCGCA ACCCAAGTCT
    401  GCACAAATTAGT AGAGGAATTG AACTCTAAGG ACTACTCCTA CCTCAATATA
    451  TTTTCTTGTG ACGGAGATAG TAGTCCTCAG CTTTTATTCC GTAAGGAATT
    501  AGAGGAAACT TCAGGGCGTT ATTTTACAGT GATTTGCGCT TTATATCTAG
    551  GGGATACAGA CATCGCTAGT TTACAACCTTG CTTCTGAAAG GATCATGGTC
    601  TCTAGAGAGT TTGATCTTGT AGATGCCTAT GCTGCAAGAT GCAAGCTCTT
    651  GAAAATCGAT CATAACAAATT GGAGACCTGG AACTTTCACT CGCACGCCG
    701  ATTTCGCAGA TGCTGTAGAC GTATCAGCAG GATTTAACTC AAGAGAATT
    751  AAAACTGATTA CGCAGGCGAA TCAAGGGATC CTAGAGTCTG GAGAACTCCC
    801  GCTCCCTTCA AAAACCTTCT GGGAGGATT CTTAGCATTC TGTGATCGAG
    851  TGACTGTAC GAGACACTTC ATTCCTTGT TAGACGCCGC TATAAAGCAA
    901  GCGGTATGGA CTCATAAAACA TCCCAAGCTTG ATAGATAAAAG AGTGTGAAGC
    951  CCTAGACTTG AAAACACAGT GCTTGCCATC TATCGTATCG TACCTTGAAT
   1001  ATGTCACAAA CTCTCACGAA AAAACATCGA AAGGCCGTT CATACAAAAA
   1051  GAGATTATCG CAGACTGTTC TCCTCTTAAA GAGGCCTCT TCCCAGGTTC

```

```

1101 TACAGAGGTG CTTGTTGAGA AAGTAACGGG GCAGGTTGCT ACGGGTCACT
1151 CTCCCTTATTG TGAAAAGGTT TCTTTCCCTG TTGTAGGAAC GGTAGCTATC
1201 AACACTCTAG TTTCTGTGCG TCTTGATAGG GTAGAGGAAG AAGGGCTGAT
1251 TGGGGAGATT GTATGA

```

5 The PSORT algorithm predicts inner membrane (0.1574).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 131A) and also as a his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 131B) and for FACS analysis.

These experiments show that cp6761 is a surface-exposed and immunoaccessible protein, and that it
10 is a useful immunogen. These properties are not evident from the sequence alone.

Example 132

The following *C.pneumoniae* protein (PID 4376766) was expressed <SEQ ID 263; cp6766>:

```

1 MATSVPVTSS TSVGEANSSN ERFTERTSRM YYAALVLGAL SCLIFIAMIV
15 51 IFPVQVGLWAV VLGFALGCLL LSLAIVFAVS GLVLGKTLEP SREATPPEIV
101 101 AQKEWTTQQD VLGNHEYWRSE LISLFLRGDL HESLIVDSKD RSLDIDQSLQ
151 NILKLEPLST TLSLLKKDCV HINIILHLVR QWNLLGVDSL PEVTAHAEEL
201 201 LLFLFLIEEQQY SPDILKLIRY GDALQATSPL MDWADSGSFS VDADGVFSCR
251 251 REECSPEDAL AQFDLLLALE NPDRRFLKDS FLTYIWSSSF FEKFLHRHLE
301 301 SLQRKLPETA IDVARYEAQI QTFLSRYFQK LDLINAMSLL WGYNCAEGEK
351 351 CYESANQRQLD NLFIIAFSSV PAMKRLFDKY GSVVVRVDRRQ IREQILSNTE
401 401 ILENESGFLC SLYEYPLSYL IDWAVLDDCV RGTEISLEDQ ADYTVCLQGL
451 451 DSMSLQSFASR LQSGQKVLPN RDVLSEQAAV MLVHGLAAQG VSFQGLKALM
501 501 YLTAVPQRMW LGALPLFESF PVFNRMKFEL GESLGD*

```

The cp6766 nucleotide sequence <SEQ ID 264> is:

```

25 1 ATGGCAACCT CTGTTCCCTGT AACTTCATCT ACTTCTGTAG GAGAGGCTAA
    51 CTCCTCCAAC GAAAGATTG CTGAACGAAC ATCGCGAATG TATTACGCAG
    101 CTTTAGTCCT AGGGGCTTTG AGCTGTTAA TTTTTTATTGC TATGATTGTC
    151 ATTTTCCCAC AGGTGCGGATT GTGGGCTGTG GTCCCTCGGGT TTGCTCTTGG
    201 ATGTTTACTT TTAAGCTTAG CTATCGTTT TGCTGTCCTC GGTCTCGTTT
    251 TAGGCAAGAC TTTAGAACCT AGTCGAGAAG CGACTCCTCC AGAAATTGTT
    301 GCGCAAAGG AGTGGACTAC ACAACAAGAT GTCTTAGGGA ATGAGTATTG
    351 GCGTTCCGAG TTGATTTCCT TGTCTTACG AGGGGATCTC CACGAATCTC
    401 TGATTGTTGA TTCTAAGGAT CGATCTTTAG ATATTGATCA GAGTTTACAA
    451 AATATATTGA AACTTGAGCC CCTATCTACG ACACTTTCGC TGTTAAAGAA
    501 AGATTGTCAC CACATCAATA TCATTTTACA TTTAGTGAGA CAGTGGAACT
    551 TACTGGGAGT GGATCTTAGT CCTGAAGTCA CTGCGCACGC CGAGGAACCT
    601 CTACTCTTT TGATAGAAGA GCAGTATTAC TCTCCTGATA TTTGAAATT
    651 GATTGCTAC GGAGATGCTT TACAAGCAAC GTCTCCTTTG ATGGATTGGG
    701 CAGATTTCAGG TTCCCTTAGT GTAGACGCAG ACGGGGTATT TAGCTGTCGC
    751 AGAGAAGAAAT GTTCTCCTGA GGATGCTTTG GCGCAATTG ATCTCTTTT
    801 GCGCITGGAA AATCCCAGACA GACGCTCTT AAAGGATCT TTTCTTACCT
    851 ACATTTGGTC GTCTTCATTT TTGAGAAGT TTTTACATCG CCATCTAGAG
    901 AGCTTGCAAA GAAAGCTCCC AGAGACAGCG ATCGATGTCG CCCGCTATGA
    951 AGCACAAATA CAAACATTT TCTCTCGCTA TTTTCAGAAG CTCGATTGA
    1001 TAAAACGCAAT GTCTCTAGAT TGCGGATATA ACTGTGCTGA GGGAGAAAAAA
    1051 TGTTATGAGA GCGCAAATCA AAGATTAGAC AACCTATTTA TTGCTTTTTC
    1101 TTCTTCTGTG CCTGCTATGA AGCGGCTCTT TGACAAATAT GGTTCTGTGG
    1151 TACGGGTAGA TCGTAGGCAG ATTCTGTGAGC AGATTCTTC GAACACTGAA
    1201 ATCTTAGAAA ATGAGTCAGG GTCTCTCTGC AGTTTGATG AATATCCTTT
    1251 ATCCTATTG ATAGATTGGG CTGTTTGCT AGACTGTGTT CGCGGTACCG
    1301 AAATCTCTCT AGAAGATCAG GCCGATTACA CGGTTGTTT GCAAGGCTTG
    1351 GATTCTATGT TATCTCAATT TGCGAGTCGT TTACAGTCTG GACAAAAAGT
    1401 ATTGAATCCT AGAGATGTTT TAAGTGAACA GGCTGCGGTT ATGCTTGTTC
    1451 ATGGCTTGGC AGCACAGGGC GTGTCGTTTC AAGGATTGAA AGCTTTGATG
    1501 TATTGACAG CCGTTCCCCA AAGAATGTGG TTAGGAGCAT TGCCCTTATT
    1551 TGAATCTTT CCTGTCCTTA ATCGGATGAA AGAATTCTT GGGGAATCTC
    1601 TGGGAGACTA G

```

5 1501 GTTATGGAAG ATGATCATAT TTTCTATGAT GTTCCTCGTA GAAAAGATGG
 1551 AATCTATGAC GTTCCTAGTT CCCCTAGATG GAGTCCTGCG CGAGAGTTGG
 1601 AAGAGGATGT TTTGGAGAT TATGAAGTTC CTATAACCTC TGCTGAACCA
 1651 TCTAAAGACA AGAACATCTA CATGACACCT AGATTAGCAA CCTCTGCTAT
 1701 CTATGATCTT CCTTCACGTC CAGGATCGTC TCCAAGCTCA CGTTCTCCGT
 1751 CTCAGATCG CGTACGAAGC AGTCACCAA ATAGACGGGG TGTGCCTCTT
 1801 CCTCCAGITC CTTCACCTGC TATGAGTGAG GAGGGGAGCA TTTATGAGGA
 1851 TATGAGCGGT GCTTCAGGTG CAGGTAAAG TGATTATGAA GATATGAGCC
 1901 GTCCCCCTC TCCTAGAGGC GACTGGATG AACCCATATA TGCTAATACT
 1951 CCTCAAGATA ATCCATTAC TCAGAGAAAT ATAGATAGAA TTTTACAGGA
 2001 GAGGTCAAGGC GGTGCTTCGG CTTCTCTGT AGAGCCTATT TATGATGAGA
 2051 TCCCCTGGAT TCATGGCAGG CCCCTGCTA CACTCCAAG ACCCGAGAAT
 2101 ACATTGACTA ATGTTTCGCT TAGAGTGAGC CCAGGGTTTC GACCAGAAGT
 2151 AAGAGCCGCT TTGCTTAGCG AGAGCGTGAG TGCTGTTATG GTCGAAGCAG
 2201 AGAGTATTGT TCCTCCAACA GAGCCGGGGG ACGGAGAAATC AGAATATCTA
 2251 GAGCCCTTAG GGGGACTTGT AGCTACAACG AAAATCTTAC TACAAAAAGG
 2301 ATGGCCTCGT GGAGAGTCGA ATGCTTAG

The PSORT algorithm predicts inner membrane (0.3994).

20 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 130A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 130B) and for FACS analysis.

These experiments show that cp6756 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 131

25 The following *C.pneumoniae* protein (PID 4376761) was expressed <SEQ ID 261; cp6761>:

1 1 MVAEVKGTF KLVCLGCRVN QYEVQAYRDQ LTILGYQEVL DSEIPADLCI
 51 51 INTCAVTASA ESSGRHAVRQ LCRQNPTAHI VVTGCLGESD KEFFASLDRQ
 101 101 CTLVSNKEKS RLIEKIFSVD TTFPEFKIHS FEKRSRAFIK VQDG CNSFCS
 151 151 YCIIPYLRGR SVSRPAEKIL AEIAGVVDQG YREVIAGIN VGDYCDGERS
 201 201 LASLIEQVDR IPGIERIRIS SIDPDDITED LHRAITSSRH TCPSSH LVLQ
 251 251 SGSNSILKRM NRKYSRGDFL DCVEKFFRASD PRYAFTTDVI VGFPGESDQD
 301 301 FEDTLRIIED VGFIKVHSFP FSARRRTKAY TFDNQIPNQV IYERKKYLAE
 351 351 VAKRVGQKEM MKRLGETTEV LVEKVTGQVA TGHSPYFEKV SFPVVGTVAI
 401 401 NTLVSVRLDR VEEEGLIGEI V*

35 The cp6761 nucleotide sequence <SEQ ID 262> is:

1 1 ATGACGGTTG CGGAAGTC AAGAACATTT AAGCTGGTCT GTPTAGGCTG
 51 51 TCGGGTGAAT CAGTATGAGG TCCAAGCATA TCGCGACCAAG TTGACTATCT
 101 101 TAGGTTACCA AGAGGTCCTG GATTCTGAAA TCCCTGCAGA TTTATGCATA
 151 151 ATCAATACGT GTGCTGTCAC AGCTTCTGCT GAGAGTTCCG GTCGTACAG
 201 201 TGTGCGTCAG TTATGTCGT AAGAACCTAC AGCACATATT GTTGTACAG
 251 251 GTTGTGTTGGG GGAATCTGAC AAAGAGTTT TTGCTCTTT GGATCGGCAA
 301 301 TGCACACTTG TTTCCAATAA AGAAAAATCC CGACTTATAG AAAAAATTTT
 351 351 TTCCCTATGAT ACGACCTTCC CTGAGTTCAA GATCCATAGT TTTGAGGGAA
 401 401 AGTCTCGAGC TTTTATTAAA GTTCAAGATG GCTGTAATTG TTTTGCTCG
 451 451 TACTGCATTA TTCTTATTG GCGGGGGCGT CGGTTTCTC GTCTGCTGA
 501 501 GAAGATTTA GCTGAAATCG CAGGGGTTGT AGACCAAGGA TATCGCGAAG
 551 551 TTGTAATTGC AGGAATTAAAT GTTGGAGATT ATTGCGATGG AGAGCGTTCA
 601 601 TTAGCCTCTT TGATTGAACA GGTGGACCGG ATTCCCTGGAA TTGAGAGGAT
 651 651 TCGAATTTC CTTATAGATC CTGATGATAT CACTGAAGAT CTGCAACCGTG
 701 701 CCATCACCTC ATCGCGTCAC ACTTGTCTT CGTCACACCT TGTTCTTCAA
 751 751 TCGGGGTCGA ATTCAATTAA AAAGAGAATG AACCGGAAGT ATTCTCGCGG
 801 801 AGATTTTTA GATTGTGAG AGAAGTTCCG TGCTTCTGAT CCTCGCTATG
 851 851 CCTTTACTAC AGATGTGATT GTCGGATTTC CTGGAGAGAG TGATCAAGAT
 901 901 TTTGAAGATA CTTTGAGAAT TATTGAAGAT GTAGGCTTTA TTAAAGTGCA
 951 951 TAGTTTCCCT TTCAGTGCTC GTCGCTGTAC TAAGGCATAT ACTTTTGATA
 1001 1001 ATCAGATTCC CAATCAGGTG ATCTATGAGA GGAAGAAGTA TCTTGCTGAG
 1051 1051 GTTGCTAAGA GGGTAGGCCA GAAAGAGATG ATGAAGCGTT TAGGAGAGAC

5 1 ATGTCATCAC TACTGAGCTG CGGAAGAATA GAGCCGACTC GGGTTACCTG
 51 TAGCTTAAAG ACGTATCTTG AGGATACGAG TCAGAACAG TTGAGCACAC
 101 GTCTAGTTCG GGCAAGTGTGTC ATCTTTTAT GCGCATTGTT GATCATTTG
 151 GTTTGTGTGG CCCTCTCTAG TTGATTCCA AGCATTATGG CCTTGGCGAC
 201 CTCTTTACG GTAATGGGGT TAATTCTTT TGTGATGTCA CTTCTGGTG
 251 ACGTTGCAAT TATAAGTTAT CTTACTTATA GCACGTGTTAC GAGTTACCGG
 301 CAAAATAAGA GAGCTTTGAG GATTCACAAG CCCGCTCGCT CCGTTTACTA
 351 CGAGGGGGTC CGCCATTGGG ATTAGGACG ATCATCTTA GGCACAGGGC
 401 AGATTCCAT AGTAAGGACG TTATTCTCTC CATTTCAGAA CCATGGTCTT
 451 AACCATGCCT TAGCTGCTAA AATTTCCTA TTTATGGAGC ATTCAGCCC
 501 TGAGCCACCG AACGAGCTT TGGTGGATTG GGCGCTGTTG ATTGGGATT
 551 TTAGGCCTCA CGTCAGTTCT TTGCTGTTG TTATGAAAA ACAAGGGTCA
 601 TCGCTGAGGA CTAAGGAAGG CAATACGATT TGTGAGGCTT TCCGCTCTGA
 651 TTACGACGCC CATTTGCTA TGGTAGATTG CTACCGGTG ATCCACTCTA
 701 AGTTGATTAGAGAAAATG GGATGAGAA ATATCGATAT CATTCCGAGT
 751 GTCATGGTTC GTGAAGAGATTA TCCTAGCCGT CCTGGGGAGG GCTATCGCGA
 801 AGGCCTATTA CGTATGTATG GTGGCAAGGG GGCTCTGTGA

The PSORT algorithm predicts inner membrane (0.711).

20 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 134A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 134B) and for FACS analysis.

These experiments show that cp6805 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 135

25 The following *C.pneumoniae* protein (PID 4376813) was expressed <SEQ ID 269; cp6813>:

1 MSGPSRTESS QVSVL SYVPR DKEIAPKKQF TIAKISTLAI LASLALGALV
 51 AGISLTIVLQ NPVLFLALLIT TALFSVVTFL VYHQMTSKVS SNWQKVLEQN
 101 FKPLGKAWQE KNVDCYSNEM QFYNNHLNPK FKVAIQTDAS QPFQPTFLTG
 151 LRVIEKNQST GIIFNPVGPT NLIDNTATNL STILYSTLKD KSVWDTCKQR
 201 EGGPAKGEDP FSPTEVRRVK LPNEALDQTF NLNLSSAEKK SILPTFLGHV
 251 CGPKSEELPN QQEYYRQALL AYENCLKAII ESHAAIVALP LFTSVYEVPP
 301 EEEILPKEGTF YWDNQTQAFK KRALLDAIQN TALRPQRSL LVILQDPFNT
 351 IESQSRSSEE*

The cp6813 nucleotide sequence <SEQ ID 270> is:

35 1 ATGTCAGGAC CCTCACGTAC TGAGAGCTCT CAAGTTCTG TACTATCCTA
 51 TGTGCCTCGG GATAAAAGAAA TTGCTCCTAA AAAACAGTT ACCATAGCAA
 101 AAATATCCAC TCTTCAACATC CTAGCTTCTT TAGCTTTAGG AGCTTTGGTG
 151 GCTGGAATCT CTTAACGAT AGTATTAGGG AACCTGTAT TTTTGGCTCT
 201 TCTCATTACC ACGGCCCTCT TCTCAGTTGT AACCTTCTTA GTCTACCACC
 251 AAATGACCTC AAAGGTATCT TCTAACTGGC AGAAAGTTCT AGAGAAAAC
 301 TTCAAGCCTT TGGAAAAGC GTGCAAGAA AAAAACGTTAG ACTGCTACTC
 351 AAACGAGATC CAATTTTACA ATAATCACCT GAACCTTAAG TTCAAGGTAG
 401 CGATACAAAC AGATGCGTCT CAACCATTTG AGCCTACTTT CTTAACTGGA
 451 CTTAGAGTGA TCGAAAAAAA TCAATCCACA GGATCATCT TTAATCCCCT
 501 AGGCCCAACG AATCTGATCG ACAACACTGC AACGAACCTC TCTACTATCC
 551 TTACTCCAC CCTAAAAGAT AAAAGCGTGT GGATACATG CAAGCAACGC
 601 GAAGGGGGTC CCGCAAAGG AGAAGACCCC TTTTCCCTA CCGAAGTGAG
 651 AGTAGTAAAA CTTCCAAACG AAGCTCTAGA TCAAACTTT AATCTAAATT
 701 TAAGCTCTGC AGAAAAGAAA AGTATTCTTC CGACCTTTT AGGCCACGTA
 751 TGCAGCCCTA AATCTGAAGA GTTACCAAT CAGCAAGAAT ATTATGCCA
 801 AGCTTTACTA GCGTACGAGA ACTGCCTTAA AGCAGCTATA GAAAGTCATG
 851 CAGCAATCGT TGCTCTCCCT CTCTTTACTT CGGTCTATGA AGTGCCTCCA
 901 GAAGAGATTC TTCCTAAAGA AGGCACCTTC TATTGGGACA ACCAAACTCA
 951 AGCGTTTGC AAACGCGCTT TATTGGACGC TATTCAAAT ACGGCCCTAC
 1001 GCTATCCTCA AAGATCTTTA CTTGTTATAC TCCAAGATCC TTTTAATACT
 1051 ATAGAATCAC AAAGTCGTTC TGAGGAGTAA

-157-

The PSORT algorithm predicts inner membrane (0.6158).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 132A) and also as a his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 132B) and for FACS analysis.

- 5 These experiments show that cp6766 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 133

The following *C.pneumoniae* protein (PID 4376804) was expressed <SEQ ID 265; cp6804>:

```

10      1  MSNQLQPCIS LGCVSYINSF PLSLQLIKRN DIRCVLAPPA DLLNLILLIEGK
          51 LDVALTSSLG AISHNLGYVP GFGIAANQRI LSVNLYAAPT FFNSPQPRIA
          101 ATLESRSSIG LLKVLCRHLW RIPTPHILRF ITTKVLRQTP ENYDGLLLIG
          151 DAALQHPVLP GFVTYDLASG WYDLTKLPFV FALLLHSTSW KEHPLPNLAM
          201 EEALQQFESS PEEVLKEAHQ HTGLPPSLLQ EYYALCQYRL GEEHYESFEK
          251 FREYYGTLQ QARL

```

- 15 The cp6804 nucleotide sequence <SEQ ID 266> is:

```

20      1  ATGTCTAACCC AACTCCAGCC ATGTATAAGC TTAGGCTGCG TAAGTTATAT
          51 TAATTCCCTT CCGCTGTCCC TACAACTCAT AAAAAGAAAC GATATTCGCT
          101 GTGTTCTTGC TCCCCCTGCA GACCCTCCTCA ACTTGCTAAT CGAAGGGAAA
          151 CTCGATGTTG CTTTGACCTC ATCCCTAGGA GCTATCTCTC ATAACCTTGGG
          201 GTATGTCCCC GGCTTTGGAA TTGCAGCAAA CCAACGTATC CTCAGTGTAA
          251 ACCCTCTATGC AGCTCCCCTC TTCTTTAACT CACCGCAACC TCGGATTGCC
          301 GCAACTTTAG AAAGTCGCTC CTCTATAGGA CTCTTAAAG TGCTTTGTCG
          351 TCATCTCTGG CGCATCCCAA CTCCCTCATAT CCTAAGATTC ATAACCTACAA
          401 AAGTACTCAG ACAAAACCCCT GAAAATTATG ATGGCCTCCT CCTAACCGGA
          451 GATGCAGCGC TACAACATCC TGTACTTCCT GGATTTGTAA CCTATGACCT
          501 TGCCTCGGGG TGGTATGATC TTACAAAGCT ACCTTTTGTA TTTGCTCTTC
          551 TTCTACACAG CACCTCTTGG AAAGAACATC CCCTACCCAA CCTTGCGATG
          601 GAAGAAGCCC TCCAACAGTT CGAATCTTCA CCCGAAGAAG TCCTTAAAGA
          651 AGCTCATCAA CATAACAGGTC TGCCCCCTTC TCTTCTTCAA GAATACTATG
          701 CCCTATGCCA GTACCGTCTA GGAGAAGAAC ACTACGAAAG CTTTGAAAAAA
          751 TTCCGGGAAT ATTATGGAAC CCTCTACCAA CAAGCCGAC TGTAA

```

The PSORT algorithm predicts inner membrane (0.060).

- The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 133A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 35 133B) and for FACS analysis.

These experiments show that cp6804 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 134

The following *C.pneumoniae* protein (PID 4376805) was expressed <SEQ ID 267; cp6805>:

```

40      1  MSSLLSCGRI EPTRVTCMSL TYLEDTSQNQ LSTRLVRAASV IFLCALLIIL
          51 VCVVALSSSLIP SIMALATSFT VMGLLIFVMS LLGDVAIIISY LTYSTVTSYR
          101 QNKRAFEIHK PARSVYYEGV RHWDLGRSSL GTGEIPIVRT LFSPFQNHLG
          151 NHALAAKIFL FMEHFSPEPP NEPLVDWACL IRDFRPHVSS LCFVIEKQGS
          201 SLRTKEGNTI CEAFRSDYDA HFAMVDCYRL IHSKLIIEKM GLKNIDIIPS
          251 VMVREDYPSR PGEGYREGLL RMYGGKGAL*

```

- The cp6805 nucleotide sequence <SEQ ID 268> is:

201 CIGFFGINGI CSTFLMLTNP RSRRDRWRNL RIMVLCYRSL GSGMNLFDLS
 251 NNRMAARRH VTSCTVALYA MVTLFGWTV A IQDALQYGF SVRDAFYRYC
 301 LRHRYCLTQR NEDSLQTTGT RFQVTRTHLE DQQMVASI LN LSVFGLFFGF
 351 VGLMTTFGGL EISPSCRWDA ANNRTVGIF*

- 5 The cp7201 nucleotide sequence <SEQ ID 274> is:

1 GTGCTCGTTG GTATCTGTCC TTCTCTATAT CCAGAACATC CTCGCTCCTT
 51 TTATTATCGT GTTTCTGGAG ATATAGGCTC CCGATTGAC GATAGAGGAT
 101 TTGTAAACTC TGGAGTCGAA ACCCTGCCAT ACTCTTCAGG CAGCTTG
 151 ATTGTTTGGA TCTCGTTTAC GGATCCCACA TTTAATTTCG CTATCGTAAA
 201 TACCTTTATG CGAACTGCAG GGATCAATGA AGTCTCTAGA CCCATGACAC
 251 AAGATACAGA AACTTCATTG ATAGAAATGA GAGACCTAAG TGAACAAACAA
 301 GAAGCGAATA ACACAGATT C TTAGAGCAA GAAGAGAGCT TAATGGGTAT
 351 TGTAGGACAT ACTGTGGGAG GAGTTCCAT GACCCTGAC TCCAGTCAA
 401 ATATCTTTA TCGTATAACAA ACACTTCTGG GACTGCCAGA GACTCTTGCA
 451 GAAGCTGAAG AAAATCCTAC CTTCCCAAAT TCTACTATAG ATAGCCTTG
 501 AGAAATAATG ATGAACCTCG TAAGGATCTC TGATGCTGTC TCTATTTCT
 551 GGATTTTCC TATCGTAGAT ACTACATATA ATGGAGTTT ATTAGCCGTC
 601 TGATCGGCT TCTTCGGAAT CAATGGGATT TGTTCCACGT TCCTTATGCT
 651 TACGAATCCA CGCTCTCGTC GAGATAGAT GAGGAATTAA CGCATCATGG
 701 TTCTTGTCA TCGTTCTTTCG GGAAGCGGAA TGAATCTCTT TGATCTTAGC
 751 AATAATGTGC GCATGGCAGC ACGTAGGCAT GTGACATCAT GTACAGTAGC
 801 TCTCTATGCT ATGGTCACTC TATTGGATG GACAGTAGCA ATACAAGATG
 851 CTTTGCATAA TGGTTTCCCT AGCGTTCGGG ATGCCTCTCA TAGATATTGC
 901 TTACGCCACA GATATTGCTT AACTCAAAGA AACGAAGACT CTCTGCAAAC
 951 TACAGGAACG CGCTTTCAGG TTACCCGTAC ACATCTAGAA GATCAACAGA
 1001 TGGTGGCTTC TATTTGAAT TTGAGTGT TTGGGCTCTT TTTGGATTC
 1051 GTAGGGCTAA TGACCACGTT TGGAGGATTA GAAATCTCAC CATCTGTGCG
 1101 GTGGGATGCA GCAAATAACC GAACGGTAGG TATTTTTAG

The PSORT algorithm predicts inner membrane (0.3102).

- 30 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 137A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 137B) and for FACS analysis.

These experiments show that cp7201 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

35 Example 138

The following *C.pneumoniae* protein (PID 4377251) was expressed <SEQ ID 275; cp7251>:

1 MAPIHGSNAF VEDILHSHPS PQATYFSSTR AQKLHEFKDR HPVLTRIASV
 51 IIKIFKVLIQG LIILPLGIW LCQLCTNSI LPSKNLLKIF KKQPNTKTLK
 40 101 TNYLHALQDY SSKNRVASM RVPILQDNVL IDTLEICLSQ APTNRWMLIS
 151 LGSDCSLEEI ACKEIFDSWQ RFAKLIGANI LVYNYPGVMS STGSSLKDL
 201 ASAAHNICTRY LKDKEQGPAG KEITIYGYSL GGLIQAEALR DQKIVANDDT
 251 TWIAVKDRCP LFISPEGFHS CRRIGKLVAR LFGWGTKAVE RSQDLPCL EI
 301 FLYPTDSLRR STVRQNLLA PELTLAHAIK NSPYVQNKEF IEVRLSSDID
 351 PIDS KTRVAL ATPILKKLS*

- 45 The cp7251 nucleotide sequence <SEQ ID 276> is:

1 ATGGCTCCAA TTCACGGAAG TAATGCGTTT GTTGAGGATA TTTTACATTC
 51 CCACCCCTCT CCACAAGCGA CTTATTTTC TTCAACACGC GCCCAAAAAC
 101 TTCACTGAGTT TAAAGACAGG CATCCCGTGC TTACACGGAT TGCTTCTGTA
 151 ATTATTAAAAA TTTTTAAAGT TCTGATAGGG CTGATCATCC TTCCCTTAGG
 201 AATCTACTGG CTATGTCAAA CGCTTGTAC AAAACTCGATT CTCCCTTCCA
 251 AGAATTTTATT AAAAATTTC AAGAAGCAAC CCAACACTAA AACCTTAAAA
 301 ACTAATTATT TGCATGCTTT GCAAGATTAT TCCTCGAAAA ACCGGCGTTGC
 351 TTCCATGAGA CGAGTTCTA TCCTCCAGGA TAATGTTCTC ATCGACACTT
 401 TGGAAATATG CCTTTCACAA GCACCTACGA ATCGTTGGAT GCTCATTTCT
 451 TTAGGAAGTG ACTGTAGCTT GGAAGAAATC GCTGTAAAGG AGATCTTTGA

The PSORT algorithm predicts inner membrane (0.4291).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 135A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 135B) and for FACS analysis.

- 5 These experiments show that cp6813 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 136

The following *C.pneumoniae* protein (PID 4376844) was expressed <SEQ ID 271; cp6844>:

10	1 MWRVVLRFI IFLIGRAVFP LRASESFSWE TSTCLTVLGI PFIDIIILTTN 51 EDFVAQCGLQ IGTISSTNNAA KIKEIFLIYK EKFPEASISF KRKEPLNLSQ 101 SHLSDLGILC MRNGETYAEQ MANKENGPAL KQPKDRLVVL RCPNQPDPLL 151 YSEKEAEKGII ETNTCLCNQG YTLLDGQLIL YGDSIEKFLK ETKRKNHNTL 201 VDLCDSQVVT TFLGRFWSSL NYVQVLFNSE DSAKILAGIP DLAQATQLLS 251 HTVPLLFIYT NDSIHIEQG KESSFTYNQD LTEPILGFLF GYINRGSMEY 301 CFNCAQSSLG ET*
----	--

The cp6844 nucleotide sequence <SEQ ID 272> is:

20	1 ATGTGGCGCG TTGTCCTCAG ATTCCCTTATA ATTTTTATCT TGGGAAGAGC 51 CGTCTTCCCT CTAAGAGCTT CAGAAAGCTT CTCTGGGAA ACATCGACCT 101 GTTTAACAGT GCTAGGGATT CCTTTCATAG ATATTATCCT CACAACGAAT 151 GAGGACTTTG TTGCCAGTG CGGCCTGCAA ATAGGAACCA TTTCTTCGAC 201 TAATAACGCA AAAATAAAAG AAATTTTTT GATATATAAG GAAAAATTTC 251 CAGAAGCCTC TATCAGTTTC AAACGAAAAG AACCTCTAAA CCTTCCCAC 301 TCCCACATCTCT CCGATTTAGG TATTTTATGT ATCGTAACG GAGAAACTTA 351 CGCTGAGGGGA ATGGCAAATA AAGAAAACGG ACCCGCTCTA AAACAACCCA 401 AGGATCTAAG ATTAGTTTTA CGTTGTCCTA ACCAACCGA TACCCGCTC 451 TACTCGGAAA AAGAACGAGA AAAGGGCATA GAAACAAATA CTTGCCTATG 501 CAATCAGGGGA TACACACTCC TGGATGGGCA ATTGATTCTC TACGGGGATA 551 GTATAGAAAA GTTTCTGAAA GAGACCAAAA GAAAGAATAA CCACACGCTT 601 GTTGATCTTT GTGACTCACA AGTCGTGACC ACGTTCCCTCG GTCGCTTTG 651 GTCTCTTCTA AACTACGTTT CAGTTCTTT CCTATCTGAA GACTCCGCTA 701 AAATTCTTGC GGGCATCCCA GACCTAGCTC AAGCTACGCA ATTGCTTTCC 751 CACACCGTAC CTPTGCTTT TATTTATACC AACGATTCTA TTCACATCAT 801 AGAACAAAGGC AAAGAAAAGTA GTTTTACCTA TAACCAAGAT TTAACAGAGC 851 CCATTTAGG ATTTCTCTTT GGTTACATAA ATCGCGGCTC TATGGAATAC 901 TGCTTTAATT GTGCACAGTC TTCATTAGGA GAAACCTAA
----	--

The PSORT algorithm predicts inner membrane (0.1786).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 136A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 136B) and for FACS analysis.

- 40 These experiments show that cp6844 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 137

The following *C.pneumoniae* protein (PID 4377201) was expressed <SEQ ID 273; cp7201>:

45	1 VLVGICPSLY PEHPRSFYYR VSGDIGSRFD DRGFVNNSGVE TLPYSSGSFG 51 IFWISFTDPT FNFAIVNTFM RTAGINEVSR PMTQDTETSL IEMRDLSEQQ 101 EANNTDSLEQ EESLMGIVGH TVGGVSMVT SSPNIFYRIQ TLLGLPETLA 151 EAEEENPTFPN STIDSLAEIM MNLVRISDAV SIFWIFFPIVD TTYNGVLLAV
----	---

1251 CGTCTTTAAA TCCATGCAAA AAGCAGATCC AGAAACCAAA GCTTTAATCC
 1301 GTGAGTTGC TCTAGATATA TTATATGCAT CCTTACGGCT TCCTCAA
 1351 TCCGCTCAT A CCGAGGTCTT TTCTACACTC TTAATGGACC CAGAGACCTA
 1401 TGAACCTAAT AAAGCTTGTA TCGCCTACTT GCTCTATGTA TTAAAGATCA
 1451 TCGAAGTATA A

5

The PSORT algorithm predicts inner membrane (0.5989).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 139A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 139B) and for FACS analysis.

- 10 These experiments show that cp7288 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 140

The following *C.pneumoniae* protein (PID 4377359) was expressed <SEQ ID 279; cp7359>:

15 1 MPGVSSSPPL SPVIVRERVP SSSGSDLIQP HAVLKISILIL FALVTLIGIV
 51 LVVLSSALGA LPSLVLTSGC CIAIAVGLIG LGILVTRLIL STIRKVDAMG
 101 YDAAVKEEQY LSRIRELESE NREIRDRNRA VEDQCAHLSE ENKDLRDPEY
 151 LHGMTERLIA SLEIENQALV AENILLKDWN ASLSRDFRAY KQKFPLGAE
 201 PWKEDIACIM EQNLFLKPEC IAMVKSLPLE TQLRLFLYPKG FQSLVNRFAP
 251 RSRFFQTPKY EYNRSRNENED GKVAAVCARL KKEFFSAVLG ACSYEELGGI
 301 CERAVALKET LPLPEAVYDT LVQEFPNLLT AESLWKEWCF YSYPYLRPYL
 351 SVDYCKRLFV QLFEELCLKL FTGSPEDQA LVRLFSYYRN HIPAVLASFG
 401 LPPPETGGSV FVLLPKQENL LWSQIEVLAT RYLKDTFVRN SEWTGSFEMM
 451 FSYNEMCKEI SEGRIRFAED YETRHSEEFP PSPLSEECEG EEEFLPPCSEE
 501 EVSVLERPDL DVDSMWVWHP PVPKGPL*

- 25 The cp7359 nucleotide sequence <SEQ ID 280> is:

30 1 ATGCCAGGTT CTGTGTCATC ACCTCCTTG TCTCCTGAA TTGTCCTGTGA
 51 AAGGGTCCCA TCCTCTTCAG GATCCGACCT CATACAGCCT CATGCTGTTT
 101 TAAAGATCTC CATCCTAATT TTTGCGCTTG TGACAATTAGGTT AGGAATTGTT
 151 CTTGTAGTGT TGTCTAGTGC TTTAGGAGCT CTTCCCTAGTT TAGTTTGAC
 201 GGTTCCTGGT TGTATTGCAA TAGCTGTAGG CCTGATTGGT TTAGGGATTTC
 251 TTGTCACACG GCTGATTCTC TCTACGATCA GAAAAGTAGA TGCCATGGGT
 301 TATGATGCTG CGGTCAAAGA AGAGCAGTAT TTGTCACGTA TCAGAGAATT
 351 AGAGTCTGAA AATAGAGAGA TTAGAGATAG AAATCGTGC GTCGAAGATC
 401 AGTGTGCCA TTTATCCGAA GAGAACAAAGG ACCTTACGGGA TCCCCAATAT
 451 CTACATGGAA TGACTGAAAG GCTCATTGCG AGCTTAGAAA TAGAGAATCA
 501 AGCTCTCGTA GCTGAGAACAA TTCTCTCAA AGACTGGAAT GCAAGCCTAT
 551 CTAGAGATTG CCGCGCATAT AACAAAAAT TTCCCTCTGG GGCATTAGAA
 601 CCCTGGAAAG AAGATATTGC ATGTATCATG GAACAAAATC TCTTTTAA
 651 ACCGGAATGT ATCGCGATGG TTAAGTCTCT TCCATTAGAG ACGCAACGGC
 701 TGTTCCTTATA TCCAAAAGGA TTTCAGTCTT TAGTTAATCG ATTGCTCCG
 751 CGCTCTCGCT TTTTCCAGAC TCCAAAAGTAT GAATATAACA GTAGGAATGA
 801 AAATGAGGAC GGAAAGGTAG CCGCAGTGTG CGCCCGTTG AAAAAGAAT
 851 TCTTCAGTGC TGTTCCTAGGA GCCTGTAGTT AGCAAGAACT AGGGGGCATT
 901 TGTGAAAGAG CAGTAGCACT TAAAGAGACG TTGCAATTGC CTGAAGCTGT
 951 CTATGATACC CTAGTTCAAG AGTCCCCAA TCTTCCTACT GCTGAGAGTT
 1001 TATGGAAAGA ATGGTGCCTTC TATTCCTATC CCTACCTTCG TCCCTATCTT
 1051 TCTGTGGATT ACTGTAAGAG GTTATTGTA CAACTTTTG AGGAACCTCG
 1101 CCTAAAGCTT TTTACAACGG GATCTCCAGA AGACCAAGCT TTGGTTCGCC
 1151 TTTTCTCTTA CTATAGGAAT CATATTCCCG CAGTCTTGGC CTCATTGGT
 1201 TTGCCCCCGC CTGAGACAGG GGGGTCTGTA TTGTTATTGC TACCAAAACA
 1251 AGAAAACCTT CTTTGAGTC AAATTGAGGT GCTGGCTACA AGGTATCTCA
 1301 AAGATACCTT CGTGAGAAC TCAGAATGGA CGGGCTCTT CGAGATGATG
 1351 TTTTCTTATA ACGAGATGTG TAAGGAGATC TCCGAAGGAA GGATTCGTTT
 1401 TGCTGAAGAC TATGAAACGA GGCATTCCGA AGAATTCCCT CCTTCCCCTC
 1451 TCTCTGAAGA AGGAGAGGGC GAAGAATTCC TTCCCTCCTG CTCTGAAGAA
 1501 GAGGTTTCGG TTCTTGAGCG CCCAGATCTA GATGTAGACT CTATGTGGGT
 1551 CTGGCATCCCG CCGGTCCCTA AGGGACCTCT TTAA

55

-161-

5 501 TTCTGGCAA AGATTTGCCA AGTGTGATAGG GGCCAATATA CTCGTTTATA
 551 ACTACCCCCG AGTCATGTCC AGCACAGGG GCAGCAGCCT AAAGGACCTA
 601 GCATCAGCTC ATAATATTG TACAAGATAC CTTAAAGATA AAGAACAGGG
 651 CCCCTGGAGCA AAAGAAAATCA TTACCTATGG GTACTCCCTA GGAGGTTTGA
 701 TACAAGCAGA AGCATTGCGA GACCAGAAGA TTGTTGCAAA CGATGATACT
 751 ACTTGGATAG CAGTCAAAGA TAGGTGTCCT CTCTTTATAT CTCCAGAAGG
 801 TTTCCACAGT TGCAGACGCA TAGGAAAGCT AGTAGCTCGT CTTTTTGGCT
 851 GGGGGACCAA AGCCGTAGAG AGAACGCAAG ACCTTCCCTG CCTAGAAAATT
 901 TTTCTCTATC CTACGGATTC CTTACGAAGA TCAACAGTCA GACAGAACAA
 951 GCTCTTAGCA CCTGAACCTA CTCTCGCTCA TGCGATAAAA AATAGTCCCT
 10 1001 ATGTTCAAAA TAAAGAATT ATAGAAGTAC GATTATCGTC TGATATCGAT
 1051 CCCATCGACA GCAAAACAAG AGTGGCTCTT GCCACACCAA TTTTGAAAAAA
 1101 GCTCTCTTAG

The PSORT algorithm predicts inner membrane (0.4545).

- 15 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 138A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 138B) and for FACS analysis.

These experiments show that cp7251 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

20 Example 139

The following *C.pneumoniae* protein (PID 4377288) was expressed <SEQ ID 277; cp7288>:

25 1 MHMSNPISLF SPAELIAKYN LIPKTSPIPP RRTELIILEE NACQTRLTNV
 51 AQVLHPSSLF SMSKKILNPC GCSGGPLCWV ILNLILAFIT SVLFILLPV
 101 NLIVAGLRLF MPLPPKKIVE DLSEPTTEET NEVIQPFI A LQALLFEDNK
 151 LRSFKIVEQS VGKAPLPNPF LNRLVAISPQ ESQEAMRKIP DLCSQLKKVL
 201 KSLGVLTPEW KHMLKYFEGL KNEHDSNPDK KTFPILIKLL IEALTGKSSL
 251 PKTPSTKEKM QAALFIASSC KTCKPTWGEV ITRSLNRLLS IANEGDNQLL
 301 IWVQEKFERE LMSIQDGDDA EEEYRFAAQHQH GERYTAIEQ VLRNESAAKL
 351 QWHVINTMKF FHGKNLGLVT EHLQDTLGL TLRQTTVDTH QGREDAADLSA
 30 401 ALFLNKYLNS GNQLVNSVFK SMQKADPETK ALIREFALDI LYASLRLPQT
 451 SAHTEVFSTL LMDPETYEPM KACIAYLLYV LKIEL*

The cp7288 nucleotide sequence <SEQ ID 278> is:

35 1 ATGCATATGT CTAACCCCAT CTCTTTGTTT TCCCCTGCAG AGTTAATAGC
 51 AAAGTACAAT TTAATTCCAA AAACCTCGCC GATTTATCCT CGGAGGACGG
 101 AACTTATTAT CTTGGAAGAA AATGCGTGTG AAACACGCT AACCAAACGTG
 151 GCTCAGGTCC TACATCCCTC TAGCCTATTG AGTATGTCAA AAAAATACT
 201 GAATCCCTG GGGTGCTCTG GTGGTCCCTT ATGTTGGGTG ATTCTCAACA
 251 TCCTAGCATT TATTATTACT TCAGTACTGT TTATCATCT TTTACCGGTG
 301 AATCTCATCG TAGCAGGTCT TCGTCTCTC ATGCCCTCTC CCCCTAAAAA
 351 AATCGTAGAG GATTTAAGTG AACCTACTAC TGAAGAAACG AATGAGGTCA
 401 TTCAACCCCTT CATTTCGCT TTGCAAGCGT TGCTTTTGA GGATAACAAA
 451 CTTCGCTCTT TTAAAATTGT TGAACAAAGT GTAGGCAAAG CACCCCTTACC
 501 TAATCCCTTT TTAAATAGAC TAGTAGCAAT TTCGCCGCAA GAAAGCCAAG
 551 AAGCCATGCG GAAGATTCCG GATCTATGCT CACAAGTAA AAAAGTATTA
 601 AAGTCTCTAG GCGTGCTAAC TCCAGAAATGG AAGCACATGC TGAAGTACTT
 651 TGAGGGACTG AAAAACGAAAC ATGATAGTAA TCCTGATAAA AAGACGTTCC
 701 CAATATTGAT CAAGCTCCTC ATAGAAGCTC TTACTGGAAA GTCCTCTTAA
 751 CCCAAAATC CTAGTACAAA GGAAAAAATG CAAGCGGCCT TATTTATTGCA
 801 AAGTTCTTGC AAGACTTGTG AGCCGACTTG GGGAGAAGTC ATAACCAGAT
 851 CTCTTAACAG ACTCTATAGT ATAGCTAATG AAGGAGACAA TCAGCTTCTG
 901 ATTTGGGTTT AAGAGTTAA AGAACGAGAG CTGATGTCCA TCCAAGATGG
 951 TGATGATGCT GAAGAGTATC GGTGTCGGC TCAGCAACAC GGTGAGCGTT
 50 1001 ACACAGAGGC AATAGAACAA GTTCTACGAA ACGAGTCAGC AGCCAAACTA
 1051 CAATGGCATG TGATCAACAC TATGAAATTG TTCCATGGGA AAAATCTCGG
 1101 TCTAGTTACA GAACACCTAC AAGATACTCT CGGCGCCCTA ACTTTACGTC
 1151 AAACTACAGT GGACACACAT CAAGGCAGAG AAGACGCTGA TTTGTCAGCT
 1201 GCTCTTTTCC TAAATAAGTA TTAAATTCT GGAAATCAAC TTGTTAATAG

Example 142

The following *C.pneumoniae* protein (PID 4377377) was expressed <SEQ ID 283; cp7377>:

```

5      1 MREETVSWSL EDIREIYHTP VFELIHKANA ILRSNFLHSE LQTCYLISIK
      51 TGGCVEDCAY CAQSSRYHTH VTPPEPMMKIV DVVERAKRAV ELGATRVCLG
     101 AAWRNAKDDR YFDRVLAMVK SITDLGAEV C CALGMLSEEQ AKKLYDAGLY
     151 AYNHNLDSSP EFYETIITTR SYEDRLNLLD VVNKGSGISTC CGGIVGMGES
     201 EEDRIKLLHV LATRDHIPES VPVNLLWPID GTPLODQPPI SFWEVLRTIA
     251 TARVVFPNSM VRLAAGRAFL TVEQQTLCFL AGANSIFYGD KLLTVENNDI
     301 DEDAEMIKLL GLIPRPSFGI ERGNPCYANN S*

```

10 The cp7377 nucleotide sequence <SEQ ID 284> is:

```

1      1 ATGCGTGAAG AAACTGTATC CTGGTCATTA GAAGACATCC GCGAAATTIA
      51 TCACACTCCC GTATTTGAGC TGATTCAAA AGCCAATGCC ATATTGCGTA
     101 GTAATTTCTT CCATTCAGAA CTGCAGACTT GCTATCTGAT TTGATTAAA
     151 ACTGGTGGAT GCGTTGAAGA TTGCGCCTAC TGTGCCAAT CTTCCGCTA
     201 TCATACCCAC GTCACACCAG AACCTATGAT GAAAATTGTA GACGTTGTGG
     251 AAAGGGCAAA ACGTGCTGTA GAGCTAGGCG CCACTCGTGT GTGCTTGGG
     301 GCTGCCTGGC GCAATGCTAA GGACGATCGA TACTTTGATA GAGTCCTCGC
     351 TATGGTGAAGA AGTATCACAG ATCTCGGAGC CGAGGTTGTG TGTGCTTAG
     401 GCATGCTCTC CGAAGAGCAA GCTAAAAAAC TGTATGATGC AGGACTTTAT
     451 GCCTACAATC ATAATTAGA CTCTTCTCCG GAATTCTATG AACTATAAT
     501 CACAACACGT TCTTATGAAAG ATGCCCTCAA CACTCTTGAT GTAGTAAATA
     551 AATCTGGCAT TAGTACATGC TGCGGTGGTA TTGTAGGTAT GGGAGAATCT
     601 GAAGAAGACC GTATAAAGCT TCTTCATGTT CTTGCAACAA GAGATCATAT
     651 CCCAGAATCC GTACCTGTA ATTACTTTG GCGGATTGAC GGCACGCCCTT
     701 TGCAAGACCA GCCTCCGATT TCTTCTGGG AAGTCTTGC AACCATAGCA
     751 ACGGCACGGG TTGTTTCCC CAGATCCATG GTACGACTTG CTGCAGGACG
     801 CGCTTTCCCTC ACAGTAGAAC AACAAACCTT ATGTTTCTA GCCGGTGCCA
     851 ACTCCATATT CTATGGAGAT AAACCTGTTGA CTGTAGAAAA CAATGATATA
     901 GATGAAGATG CTGAAATGAT CAAACTTTA GGCTTAATCC CTCGCCCTTC
     951 ATTGGAATA GAAAGAGGTA ACCCATGTTA TGCCAACAAAT TCCTAA

```

The PSORT algorithm predicts cytoplasm (0.2926).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 142A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 142B) and for FACS analysis.

35 These experiments show that cp7377 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 143

The following *C.pneumoniae* protein (PID 4377407) was expressed <SEQ ID 285; cp7407>:

```

40     1 MVCNNNSWFR MCGNFNCEWV EVTTTEETTR QSASDISEEA GSSGGAAPIT
      51 TQPTKITKVE KRVQFNNTAQG DESTIHMIE AGEVLDLSILS HRRTQGCTEY
     101 CYDSYATGCG QRCGSFGR LI CGTYKACCLD REDNQVAGLV HECEQTHGPI
     151 AVALAAKTMG LNLMELVEKN TILSEEQKNE FRQHCSEAKT QLYGTMQSLS
     201 QNFFLEGVNS IRERGLDDSL VQAVLFSIAT RSWEKTISE EASGTSSASN
     251 STRIPACYIL NTSPLTTSR SCGRSDARRP SSVGAEPOYV AKKYNDNGMA
     301 RQLGKIQVTN LKTGDFSA LG PFGLLLIVKML NSFLLSASQS TSSILKHTGG
     351 EICYTCPNFR DIVVLLMLAI GYCPANTDET SVVDIHMIDD PIMTIFYRLQ
     401 YSYRTGKTS A SFLKKKPSLV RQESLDCPTP AESVPLMSSL EEEDENEDDD
     451 EDGNLAYQQR ILECSGHLQT LFLGIKINKE *

```

The cp7407 nucleotide sequence <SEQ ID 286> is:

```

50     1 ATGGTTTGCC CAAATAATC TTGGTTCAAGA ATGTGTGGAA ATTTCAACTG
      51 CGAATGGTT GAAGTAACAA CAACAGAAGA AACAAACGCGG CAATCGGCTT
     101 CAGATATAAG CGAAGAAAGCT GGTCGAGTG GAGGAGCTGC TCCTATAACT
     151 ACGCAACCTA CTAAAATTAC AAAAGTAGAG AAACGTGTCC AATTAAATAC

```

The PSORT algorithm predicts inner membrane (0.7453).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 140A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 140B) and for FACS analysis.

- 5 These experiments show that cp7359 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 141

The following *C.pneumoniae* protein (PID 4377374) was expressed <SEQ ID 281; cp7374>:

```

10      1 MDKQSSGNSG CIWHPFTQSA LDSTPIKIVR GEGAYLYAES GTRYLDAISS
      51 WWCNLHGHGH PYITKKLCEQ AQKLEHVIFA NFTHEPALEL VSKLAPLLPE
     101 GLERFFFSDN GSTSIEIAMK IAVQYYNNQN KAKSHFVGLS NAYHGDTFGA
     151 MSIAGTSPTT VPFDLFLPS STIAAPYYGK EELAIAQAKT VFSESNIAAF
     201 IYEPLLQGAG GMLMYNPEGI KEILKLAKHY GVLCIADEIL TGFGRGPLF
     251 ASEFTDIPPD IIICLSKGLTG GYLPLALTWT TKEIHDAFVS QDRMKALLHG
     301 HTFTGNPLGC SAALASLDLT LSPECLQQQRQ MIERCHQEFO EAH GSLWQRC
     351 EVLGTVLAID YPAEATGYFS QYRDHLNRFF LERGVLLRPL GNTLYVLPPY
     401 CIQEEDLRII YSHLQDALCL QPQ*

```

The cp7374 nucleotide sequence <SEQ ID 282> is:

```

20      1 ATGGACAAGC AATCATCAGG GAATTCAAGGG TGTATCTGGC ACCCCTTCAC
      51 TCAATCTGCA TTAGATTCTA CACCCATAAA GATTGTAAGG GGAGAAGGTG
     101 CTTACCTCTA TGCGGAATCA GGAACAAGAT ATCTTGATGC GATATCTTCA
     151 TGGTGGTGC ACCTCCACGG TCATGGGCAT CCCTACATTA CAAAAAAATT
     201 ATGTGAGCAA GCACAGAACT TAGAACATGT GATCTTCGCA AATTTCACCC
     251 ATGAACCGGC TCTAGAGCTC GTATCGAAAC TCGCTCCCT CCTTCCTGAA
     301 GGTCTAGAAC GTTTCTTTT CTCTGACAAC GGATCAACGT CTATCGAAAT
     351 AGCAATGAAA ATTGCTGTGC AATATTACTA CAATCAAAAC AAGGCTAAGA
     401 GCCATTTTGT TGGACTCAGC AATGCCATTC ACGGAGATAC ATTTGGAGCT
     451 ATGTCGATAG CTGGCACGAG CCCTACTACA GTTCCCTTTC ATGATCTTTT
     501 TCTTCCTTCC AGTACAATTG CTGCTCCCTA TTATGGCAAG GAAGAGCTTG
     551 CCATTGCCCA AGCAAAACAA GTCTTTCTG AAAGCAATAT CGCAGCGTTT
     601 ATCTATGAGC CGCTTATGCA AGGTGCTGGA GGGATGTAA TGTATAATCC
     651 CGAAGGCCCTA AAGGAGATTC TCAAGCTTGC CAAGCATCAC GGGGTTCTCT
     701 GTATTGCTGA TGAAATCTT ACTGGCTTTG GCGGTACGGG TCCACTGTTT
     751 GCTTCTGAAT TTACAGACAT TCCTCCTGAC ATTATCTGTC TTTCTAAAGG
     801 TCTTACAGGA GGCTATCTCC CTCTAGCCTT GACAGTAACC ACTAAAGAAA
     851 TTCATGATGC CTTTGTCTCC CAAGATCGGA TGAAGGCCT GCTTCATGGC
     901 CATACTTCA CAGGAATCC TTTAGGCTGT AGTGTGCC CTCGCTTCTTT
     951 GGATCTCACC CTATCTCCAG AATGCCATCA ACAAAAGGCAA ATGATAGAAC
    1001 GGTGTCTACA AGAGTTCAA GAAGCTCATG GTTCCCTATG GCAACGGTGT
    1051 GAGGTTCTGG GCACGGTACT CGCTCTAGAT TACCCCTGCAG AAGCTACAGG
    1101 ATATTTTCTA CAATATAGAG ACCATCTCAA TCGCTTTTTC TTAGAACGTG
    1151 GAGTCCTTCTC TCGTCCTTAA GGGAACACAC TGTATGTGCT GCCCCCCTAC
    1201 TGTATCCAAG AAGAAGATCT CCGGATTATT TATTCTCACC TACAGGATGC
    1251 CCTATGTCTA CAACACAGT AA

```

- 45 The PSORT algorithm predicts cytoplasm (0.2930).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 141A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 141B) and for FACS analysis.

- These experiments show that cp7374 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

These experiments show that cp6432 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 145

The following *C.pneumoniae* protein (PID 4376433) was expressed <SEQ ID 289; cp6433>:

```

5   1 MNWVPKTIDH VDPESEIDIR KVVSCYKLIK ECQPEFRSLI SELLGVIRCG
    51 LRLLKRSKYQ EQARTVSDED APLFCLTRSY YQDGYLTPLR AGPRDLINHY
   101 IHLRRRENPK HFFSPKHPCY YARLAFNESV CVYRELFDIE RLTKMYVEGD
   151 YSKEQEKNLQ AILSFVKTLG EGKDFLIEHK DTDLIGRGFT DVFCT*

```

The cp6433 nucleotide sequence <SEQ ID 290> is:

```

10  1 ATGAATTGGG TTCCAAAAAC AATAGACCAT GTAGATCCAG AATCAGAGAT
    51 AGATATACTG AAAGTCGCT CCTGCTATAA GTTGATAAAA GAATGTCAAC
   101 CTGAATTTCG ATCTCTTATA AGTGAATTAC TAGGAGTGAT TCGGTGTGGC
   151 TTAAGACTAT TAAACGTTA TAAGTATCAA GAACAGGCTA GAACTGTATC
   201 TGATGAAGAT GCACCTCTTT TCTGCCTGAC TCGTTCTTAT TATCAAGATG
   251 GTTATCTCAC GCCATTAAGA GCAGGACCTC GTGATCTTAT AAATCACTAT
   301 ATACACTTGC GTCGCCGAGA GAATCCTAAG CATTTCCTCA GTCCTAAGCA
   351 TCCATGTTAT TATGCTCGAT TGGCTTTAA TGAGTCAGTG TGTGTCTATA
   401 GAGAACTCTT TGATATAGAG CGACTTACAA AAATGTATGT CGAGGGTGAT
   451 TATTCTAAAG ACAAAGAGAA AACCTACAG GCTATTCTTA GTTTGTGAA
   501 AACTCTAGAT GAAGGAAAGG ACTTTCTTAT TGAACATAAA GATACCGATC
   551 TCATTGGGAG AGGTTTACT GATGTGTTCT GCACTTAA

```

The PSORT algorithm predicts cytoplasm (0.4068).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 145A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 145B) and for FACS analysis.

These experiments show that cp6433 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 146

The following *C.pneumoniae* protein (PID 4376643) was expressed <SEQ ID 291; cp6643>:

```

30  1 MGYLPVSATD VLFESPAAPL INSANTQNQK LIELKGKQQA ESSPRTITSV
    51 ILEVLLVIGC CLIVLSSLAI RPALQFTLET GHPAIAVLA VSGTILLVAV
   101 IIIFCFCLAIV PFAAKKTYKY VKTVDDYASW HSHQQPTPLG TIFSGIVYAE
   151 SQAQL*

```

The cp6643 nucleotide sequence <SEQ ID 292> is:

```

35  1 ATGGGATATC TTCCAGTATC TGCTACGGAC GTTCTTTTG AAAGTCCAGC
    51 CGCTCCCTTA ATCAATAGCG CAAACACACA AAATCAGAAA CTCATAGAAC
   101 TCAAGGGGAA GCAGCAAGCT GAGTCTCTC CACGGACAAT CACTTCTGTC
   151 ATATTGGAAG TTCTCCTAGT GATCGGATGC TGCCCTCATAG TTCTTAGTTT
   201 ATTGGCAATC CGCCCTGCTC TGCAATTACAC TCTAGAAACT GGACATCCAG
   251 CTGCCATTGC AGTCCTTGCT GTCTCAGGAA CAATTCTATT GGTGGCTGTT
   301 ATCATCTTGT TTTGCTTCT AGCAGCTGTG CCATTCGCTG CTAAGAAAAC
   351 TTATAAAATAT GTTAAGACGG TTGATGACTA TGCTTCTTGG CATTCTCATC
   401 AGCAAACACC GACCCTAGGC ACTATCTTTT CAGGTATCGT CTATGCAGAA
   451 TCCCAGGCAGC AATTATAG

```

45 The PSORT algorithm predicts inner membrane (0.6859).

5
10
15
20
25

```

201 TGCTCAAGGT GATGAAAGTA CAATACACAT GATCCAAGAA GCAGGAGAAT
251 TGGTAGACTC CATTCTATCA CATAGACGAA CGCAAGGATG TACAGAGTAT
301 TGTTATGACA GTTACGCAAC TGGATGTGGT CAGCGTTGCG GATCTTTGG
351 AAGACTCATT TGTGGAACGT ATAAAGCGTG TTGCTTAGAC AGAGAGGATA
401 ATCAGGTTGC TGGACTTGT CATGAATGCG AACAGACCCA TGGTCCATT
451 GCCGTTGCTT TAGCTGCTAA AACTATGGC CTCAACTTAA TGGAACTTGT
501 AGAAAAAAAC ACTATTITGT CTGAAGAACAA GAAAAATGAA TTTAGACAGC
551 ATTGCTCGGA AGCTAAAACC CAACTCTATG GAACGATGCA GAGCCTTTCT
601 CAAAACCTTT TCCTTGAAGG AGTCAACAGC ATTAGAGAAC GCGGTCTAGA
651 CGATTCACTA GTCCAAGCCG TGCTAAGCTT TATTGCTACA AGGTCTTGGG
701 AAAAATCAT AGAATCAGAC GAAGCCTCAG GAACATCTTC TGCTCTAAT
751 TCTACACGCA TTCCCTGCGT CTATATCTTA AATACGAGCC CCTTAACGAC
801 GTCACGCCA TCCTGTGGAT CAAGAGATGC GCGACGCCA TCTTCAGTCG
851 GTGCAGAGCC CCAGTACGTA GCAAAAAAAT ACAATGACAA TGGCATGGCC
901 AGACAATTAG GAAAAATCCA AGTCACCAAT CTAAAAACAG GAGATTTTC
951 AGCTTTAGGT CCTTTGGTC TCCTGATTGT GAAAATGCTG AATAGCTTTC
1001 TCTTATCTGC ATCACAAAGC ACATCTCTA TTCTAAAGCA CACAGGTGGA
1051 GAAATATGTT ATACGTGCCA AAATTITCGT GATATGCTG TTTTATTGAT
1101 GTTAGCGATT GGCTATTGCC CTGCAAATAC CGATGAGACA TCTGTCGTAG
1151 ATATAACACAT GATAGATGAT CCGATTATGA CCATCTCTA TCGACTACAA
1201 TACAGCTATA GAACAGGGAA AACTTCAGCA TCGTTTTAA AAAAGAAACC
1251 CTCATTAGTA AGACAGGGAA GTCTTGATTG TCCTACCCCT GCAGAATCTG
1301 TCCCTCTCAT GTCAAGTCTC GAAGAAGAAG ATGAAAATGA AGATGATGAT
1351 GAGGATGGGA ATTTGGCGTA TCAACAGCGT ATCCTTGAAT GCTCGGGTCA
1401 TTACAAACT CTATTTTAG GGATAAAAAT AAACAAAGAA TAA

```

The PSORT algorithm predicts inner membrane (0.1319).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 143A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 143B) and for FACS analysis.

30 These experiments show that cp7407 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone:

Example 144

The following *C.pneumoniae* protein (PID 4376432) was expressed <SEQ ID 287; cp6432>:

35
51
101
151

```

1 MTRSTIESSD SLCSRFSQK LSVQTLKNLC ESRLMKITSV VIAFLTLIVG
51 GALIALAGGG VLSFPLGLIL GSVLVLFSSI YLVSCCKFT LKEMTMTCV
101 KSKINIWFKEK QRNKDIKCAL ENPDLFGENK RNVGNRSARN QLEMILHETD
151 GIILKRYMKG AKMYFYL*

```

The cp6432 nucleotide sequence <SEQ ID 288> is:

40
45
50

```

1 ATGACTAGAA GTACTATTGA AAGCAGTGAT TCGCTATGCT CAAGGTCTTT
51 TTCTCAAAAAA TTAAGTGTCC AGACATTAAA AAATCTCTGT GAAAGTAGAT
101 TAATGAAGAT CACTTCTCTT GTGATTGCTT TCCTAACTCT AATTGTTGGGG
151 GGTGCTCTTA TAGCTTTAGC AGGAGGGGGG GTTCTTTCTT TCCCTCTTGG
201 GCTAATCTTA GGAAGCGTAC TCGTTTGTT TTCTCTATC TATTTAGTCT
251 CTTGTTGTAATTTTACT TTAAAAGAGA TGACAATGAC CTGTAGTGTC
301 AAATCTAAAAA TCAATATATG GTTGAAAAG CAACGAAACA AAGACATCGA
351 AAAGGCATTA GAGAATCCAG ATCTCTTTGG AGAAAATAAG AGAAAATGTTG
401 GAAATCGTTC GGCAAGAAAT CAACTAGAAA TGATCTTACA CGAGACTGAC
451 GGAATTATTT TGAAAAGATA TATGAAAGGA GCTAAAATGT ACTTTTATTT
501 ATGA

```

50 The PSORT algorithm predicts inner membrane (0.5394).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 144A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 144B) and for FACS analysis.

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 148A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 148B) and for FACS analysis.

These experiments show that cp7253 is a surface-exposed and immunoaccessible protein, and that it
5 is a useful immunogen. These properties are not evident from the sequence alone.

Example 149

The following *C.pneumoniae* protein (PID 4376264) was expressed <SEQ ID 297; cp6264>:

10	1 VISGLLFLLV RREVPTVRSE EIPRGVSVT P SEEPALEKAQ KEPETKKILD
	51 RLPKELDQLD TYIQEVFA CL ERLKDPKYED RGLLTEAKEK LRVFDVVEKD
	101 MMSEFLDIQR VLNEEAYYVE HCQDPLENIA YEIFSSQELR DYYCAGVCGY
	151 LPSGDARAD R LKRSVKEVMD RFMRVTWKS W EASVMLDH SY GVARELFKKA
	201 VGVL EESV YK ILFKSYRDAF YECEKAKIQR DGRFKWL*

The cp6264 nucleotide sequence <SEQ ID 298> is:

15	1 GTGATTTCCGG GACTTCTATT CCTTCTAGTA AGACGAGAGG TTCCGACAGT
	51 ACGTTCA GAGAG AAAATTCCCA GAGGGGTTTC TGTGACCCCT TCTGAAGAGC
	101 CTGCTCTAGA GAAGGC TCAA AAAGAACCGG AGACAAAGAA AATTTAGAT
	151 CGGTTGCCGA AGGAATTGGA TCAGTTAGAT ACGTATATTC AGGAAGTGTT
	201 TGCATGTTA GAGAGGCTGA AGGATCCTAA GTACGAAGAT CGAGGTCTTT
	251 TAACAGAGGC GAAGGAGAAA CTTCGAGTT TGACGTTGT TGAGAAAGAT
	301 ATGATGTCAG AGTTTTTAGA CATA CAAACGA GTGTTGAATG AGGAAGCATA
	351 TTATGTAGAA CATTGTCAG ATCCCCTAGA GAATATAGCC TACGAGATTT
	401 TCTCTTCCCA AGAGCTTCGT GATTA CACT GTGCAGGGGT GTGTGGGTAT
	451 TTGCCTTCTG GGGATGCTCG AGCGGATCGA TAAAGAGAT CAGTTAAGGA
	501 GGTAAATGGAT CGCTTTATGA GGGTGACCTG GAAATCTTGG GAGGCATCAG
	551 TCATGTTGG A TCATAGCTAT GGGGTAGCGC GAGAGTTATT CAAGAAGGCA
	601 GTAGGAGTAC TAGAGGAGAG TGTCTATAAA ATTCTGTTA AGAGCTATAG
	651 AGATGCGTTT TATGAATGTG AGAAGGCAA GATCCAGAGG GATGGCGTT
	701 TCAAATGGTT ATAG

The PSORT algorithm predicts cytoplasm (0.2817).

30 The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 149A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 149B) and for FACS analysis.

These experiments show that cp6264 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 150

The following *C.pneumoniae* protein (PID 4376266) was expressed <SEQ ID 299; cp6266>:

40	1 MLLLISGALF LTLCGIPGLSA AISFGLGIGL SALGGVLMIS GLLCLLVKRE
	51 IPTVRPEEIP EGVSLAPSEE PALQAAQKTL AQLPKELDQL DTDIQEVFAC
	101 LRKLKD SKYE SRSFLNDAKK ELRVFDVVE DTLSEIFELR QIVAQEGWDL
	151 NFLINGGRSL MMTAESES LDFHVS KRLGY LPSGDV RGEG LKKSAKEIVA
	201 RLMSLHCEIH KVAVAFDRNS YAMA EKAFAK ALGALEESVY RSLTQSYRDK
	251 FLESERAKIP WNGHITWLRD DAKSGCAEKK LGMP RVGRN LGKQSFG*

The cp6266 nucleotide sequence <SEQ ID 300> is:

45	1 ATGCTCTTAC TGATTTCA GG AGCTCTCTT CTGACGTTAG GGATTCCAGG
	51 ATTGAGTGC A GCAATTCTT TTGGATTAGG CATCGGTCTC TCCCGCATTAG
	101 GAGGAGTGCT GATGATTTCG GGACTACTAT GTCTTTAGT AAAACGAGAG
	151 ATTCCGACAG TACGACCAGA AGAAATTCCCT GAAGGGTTT CGCTGGCTCC

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 146A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 146B) and for FACS analysis.

These experiments show that cp6643 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 147

The following *C.pneumoniae* protein (PID 4376722) was expressed <SEQ ID 293; cp6722>:

10	1 VSSTLNGVFP SSLPEESADL FITNKEIVAL GEKGNVFLTH SIPMHIAAIT
	51 ILVIVALAGI AIIICLGCVSQ SILLIAVGIV LTILTLCLQ ALVGFIKFIR
	101 QLPQQLHTTV QFIREKIRPE SSLQLVTNAQ RKTTQDTLKL YEELCDLSQK
	151 EFKLQSTLYQ KRFELSHKNE KTNQN*

The cp6722 nucleotide sequence <SEQ ID 294> is:

15	1 GTGTCTAGTA CTTTAAACGG GGTATTTCCC TCATCCCTTC CGGAAGAGTC
	51 TGCTGATTTA TTCATTACGA ATAAGGAGAT CGTAGCTTTG GGGGAGAAGG
	101 GCAATGTTTT TCTCACCCAC TCCATTCCCTA TGCATATTGC TGCGATTACG
	151 ATCTTAGTGA TTGTAGCTCT TGCTGGAATC GCTATTATCT GTTTGGTTG
	201 CTATAGCCAA AGCATTCTGT TGATTGCCGT TGGCATTTGTT CTTACTATTT
	251 TGACTCTTCT CTGCCTACAA GCCTTGGTAG GATTTATTAA ATTCACTCCGG
20	301 CAGCTCCCTC AGCAGCTCCA TACGACAGTA CAATTTATCA GGGAGAAGAT
	351 TCGACCTGAA TCCTCTCTAC AGCTTGTAAC CAATGCCACAG AGAAAAAACCA
	401 CTCAAAGATAC GCTAAAGTTA TACGAAGAAC TCTGCGACCT CTCACAAAAAA
	451 GAGTTCAAAAC TCCAATCAAC TCTTATCAA AAACGTTTTG AGCTTCTCA
	501 CAAGAATGAA AAGACAAATC AAAACTAG

The PSORT algorithm predicts inner membrane (0.6668).

25 The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 147A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 147B) and for FACS analysis.

These experiments show that cp6722 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

30 Example 148

The following *C.pneumoniae* protein (PID 4377253) was expressed <SEQ ID 295; cp7253>:

35	1 MSELAPCSTG LQMVPHTQVH HALDTRRVIL TIAACLSLIA GIVLVGLGAA
	51 AILPLSLFGVI GGMILILFSS IALIYLYKKK REVHQIALEP LPREMISKDQS
	101 IIDFVKTRDY ASLEKKATFA YTHHTHYDGS MVFYREIPRF MLGSYLALRK
	151 DMDRQALF*

The cp7253 nucleotide sequence <SEQ ID 296> is:

40	1 ATGAGCGAGC TCGCCCCCTG CTCGACAGGA TTGCAGATGG TCCCCCATAC
	51 GCAGGTCCAT CATGCCCTTG ATACCGGGAG AGTCATTCTA ACGATAAGCCG
	101 CCTGTCTGTC TTTAATTGCA GGAATCGTGT TGGTTGGCTT AGGTGCTGCA
	151 GCAATCCTGC CCTCGCTTT TGGAGTCATT GGAGGAATGA TTCTTATTCT
	201 GTTTTCTTCG ATGCCCTCA TTTATTTATA CAAGAAAGACA AGGGAGGTGG
	251 ATCAGATTGC TCTGGAGCCT CTTCTGAGA TGATTTCTAA AGATCAAAGC
	301 ATTATAGATT TTGTAAGAC ACGAGACTAT GCATCTTTAG AAAAGAAAGC
45	351 GACCTTTGCT TATACTCATA CTCATTATTA CGATGGAAGC ATGGTCTTCT
	401 ATAGGGAGAT CCCTAGATTT ATGTTAGGCT CTTATCTCGC GCTTCGCAAA
	451 GACATGGACC GCCAAGCTCT TTTTGA

The PSORT algorithm predicts inner membrane (0.5394).

The cp6282 nucleotide sequence <SEQ ID 304> is:

```

5   1 ATGTCCTTAT TGAACCTTCC CTCAAGCCAG GATTCTGCAT CTGAGGACTC
    51 CACATCGCAA TCTCAAATCT TCGATCCCCT TAGAAATCGG GAGTTAGTTT
    101 101 CTACTCCCAGA AGAAAAAGTC CGCCAAAGGT TGCTCTCCTT CCTAATGCAT
    151 151 AAGCTGAAC ACCCTAAAGAA ACTCATCATC ATAGAAAAG AACTCAAAAC
    201 201 TCTTTTCCT CTGCTTATGC GTAAAGGAAC CCTAATCCCA AAACGCCGCC
    251 251 CAGATATTCT CATCATCACT CCCCCCACAT ACACAGACGC ACAGGGAAAC
    301 301 ACTCACAACC TAGGCGACCC AAAACCCCTG CTACTTATCG AATGTAAGGC
    351 351 CTTAGCCGTA AACCAAATG CACTCAAACA ACTCCTTAGC TATAACTACT
    401 401 CTATCGGAGC CACCTGCATT GCTATGGCAG GGAAACACTC TCAAGTGTCA
    451 451 GCTCTCTTCA ATCCAAAAC ACAAACTCTT GATTTTTATC CTGGCCTCCC
    501 501 AGAGTATTCC CAACTCCTAA ACTACTTAT TTCTTTAAC TTATAG

```

The PSORT algorithm predicts cytoplasm (0.362).

The following *C.pneumoniae* protein (PID 4377373) was also expressed <SEQ ID 305; cp7373>:

```

15  1 MSTTTVKHFI HTASRWEVPL KEIVASNYWH AQWINTLSFL ENSGAKKISA
    51 SEHPTEVKEE VLKHAEEFRR HGHYLKTQIS RISETSLPDY TSKNLLGGGLL
    101 101 TKYYLHLLDL RTCRVLENEY SLSGQTLKTA AYILVTVYIE LRASELYPLY
    151 151 HDILKEAQSK ITVKSIIILEE QGHLQEMERE LKDLPHGEEL LGYACQFEGE
    201 201 LCLQFVERLE QMIFDPSSSTF TKF*

```

20 The cp7373 nucleotide sequence <SEQ ID 306> is:

```

1  1 ATGTCTACAA CCACAGTAAA ACACTTTATC CACACAGCCT CTCGTTGGGA
    51 GCCCGTTCTC AAAGAGATCG TAGCTTCCAA CTATTGGCAT GCACAATGGA
    101 101 TAAATACCCCT GTCCCTTTA GAAAATAGTG GAGCAAAAAA AATCTCCGCA
    151 151 AGTGAACATC CTACGGAGGT AAAGGAAGAA GTTTTAAAC ATGCTGCTGA
    201 201 AGAAATTCTCGT CATGGTCACT ATCTAAAAAC TCAGATTCT AGAATCTCAG
    251 251 AGACTTCTCT CCCTGACTAT ACATCTAAAA ATCTTCTGGG AGGCTTACTT
    301 301 ACAAAATATT ACCTCCATCT TCTAGATTAA AGGACGTGCC GAGTACTGGA
    351 351 AAATGAATAAC TCCCTATCGG GACAAACGTT AAAAATCGCA GCGTATATTT
    401 401 TAGTTACCTA CGCAATCGAA CTTCGTGCTT CTGAACCTTA TCCTCTGTAT
    451 451 CACGATATTC TGAAAGAAGC TCAAAGTAAA ATAACGGTAA AATCCATTAT
    501 501 CTTAGAAGAG CAAGGCCATC TGCAAGAGAT GGAACGTGAA CTTAAAGATC
    551 551 TCCCCCACGG GGAGGAACTC TTAGGCTATG CTTGCCAATT CGAAGGGGAG
    601 601 CTTTGCTTGC AGTTTGTAGA GAGATTAGAA CAAATGATCT TCGATCCTTC
    651 651 CTCGACTTTT ACAAAAGTTCT AG

```

35 The PSORT algorithm predicts cytoplasm (0.1069).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 152A; 6282 = lanes 8 & 9; 7373 = lanes 2-4). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 152B & 153) and for FACS analysis.

These experiments show that cp6282 & cp7373 are surface-exposed and immunoaccessible proteins 40 and that they are useful immunogens. These properties are not evident from the sequence alone.

Example 154 ,

Example 155 ,

Example 156 ,

Example 157 and

45 Example 158

The following *C.pneumoniae* protein (PID 4376412) was expressed <SEQ ID 307; cp6412>:

```

1  1 MSSSEVVVFQT VHGLGFGLS SKSVVPFKKS LSDAPRVVCS ILVLTGLGLA
    51 51 LVCGIAITCW CVPGVILMGG ICAIVLGAIS LALSLFWLWG LFSNCCGSKR
    101 101 VLPGEGLLRD KLLDGGFSRA APSGMGLPGD GSPrASTPSC LEELQAEIQA
    151 151 VTQAIIDQMSD D*

```

50 The cp6412 nucleotide sequence <SEQ ID 308> is:

201 TTCTGAGGAG CCAGCTCTAC AGGCAGCTCA GAAGACTTTA GCTCAGCTGC
 251 CTAAGGAATT GGATCAGTTA GATACAGATA TTCAGGAAGT GTTCGCATGT
 301 TTAAGAAAGC TGAAAGATTG TAAGTATGAA AGTCGAAGTT TTTTAAACGA
 351 TGCTAAGAAG GAGCTTCGAG TTTTGACTT TGTGGTTGAG GATACCCCTCT
 401 CGGAGATTTT CGAGTTGCGG CAGATTGTGG CTCAAGAGGG ATGGGATTAA
 451 AACTTTTGA TCAATGGGG ACAGAACCTC ATGATGACTG CAGAATCTGA
 501 ATCGCTTGAT TTGTTTCATG TATCGAAGCG GCTAGGGTAT TTACCTTCIG
 551 GGGATGTTG AGGGGAGGGG TTAAAGAAAT CTGCGAAGGA GATAGTCGCT
 601 CGTTTGATGA GCTTGCATTG CGAGATTCA C AAGGTGGCGG TAGCGTTTGA
 651 TAGGAATTCC TATGCCATGG CAGAAAAGGC GTTTCGAAA GCGTTGGAG
 701 CTTTAAAGA GAGTGTGTAT CGGAGTCTGA CGCAGAGTTA TAGAGATAAA
 751 TTTTGGAGA GCGAGAGGGC AAAGATCCC TGGAAATGGC ATATAACCTG
 801 GTTAAGAGAT GATGCGAAGA GTGGGTGTGC TGAAAAGAAG CTCGGGATGC
 851 CGAGGAACGT TGGAAAGAAAT TTAGGAAAGC AGTCTTTGG GTAG

15 The PSORT algorithm predicts inner membrane (0.3590).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 150A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 150) and for FACS analysis.

These experiments show that cp6266 is a surface-exposed and immunoaccessible protein and that 20 they it is a useful immunogen. These properties are not evident from the sequence alone.

Example 151

The following *C.pneumoniae* protein (PID 4376895) was expressed <SEQ ID 301; cp6895>:

1 MKIKKSFQYS LCQAKRFQNM LPNHFDPCLQ PVNLQLKQDR LAYGELIILL
 25 51 SKYQQKTFSS LLKEETCSLN RAKQHLLYKI LRDFNTMQHL RSLGLNGWGE
 101 101 IPMSPCL*

The cp6895 nucleotide sequence <SEQ ID 302> is:

1 ATGAAGATTA AAAATCTTT TCAATACAGT TTATGCCAAG CAAAGAGATT
 51 TCAGAACATG CTGCCAAACC ACTTTGATCC ATGTTTGAG CCAGTGAATT
 30 101 TACAACCAA ACAAGACAGA TTGGCATACTG GGGAGCTCAT CATATTGCTA
 151 TCTAAATATC AACAAAAGAC CTTTCCCTCT TTGTTGAAGG AAGAAACATG
 201 TTCTCTTAAT CGTGCAGAAGC AGCACTTATT GTATAAGATT TTGAGAGATT
 251 TTAATACTAT GCAGCATCTA AGGTCCCTCG GATTAAATGG TTGGGGAGAG
 301 ATCCCTATGA GTCCTTGCT CTAA

The PSORT algorithm predicts cytoplasm (0.3264).

35 The protein was expressed in *E.coli* and purified as a his-tag product (Figure 151A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 151B) and for FACS analysis.

These experiments show that cp6895 is a surface-exposed and immunoaccessible protein and that it is a useful immunogen. These properties are not evident from the sequence alone.

40 **Example 152 and**
Example 153

The following *C.pneumoniae* protein (PID 4376282) was expressed <SEQ ID 303; cp6282>:

1 MSLLNLPSQQ DSASEDSTSQ SQIFDPIRNR ELVSTPEEKV RQRLLSFLMH
 25 51 KLNYPKKLII IEKELKTLFP LLMRKGTLP KRRPDILIIIT PPTYTDAQGN
 101 101 THNLGDPKPL LLIECKALAV NQNALKQLLS YNYSIGATCI AMAGKHSQVS
 151 151 ALFPNPKTQTL DFYPGLPEYS QLLNYFISLN L*

The PSORT algorithm predicts inner membrane (0.5989).

The following *C.pneumoniae* protein (PID 4376654) was also expressed <SEQ ID 315; cp6654>:

```

5      1 MKTKMNSRKK AGQWAIFNSP TPVGSSTLVL AWTPWGYYDK DVQDILERKD
      51 PMSSSLSEKD SKEFLKNLFV DLLENGFTSV HIHAEEAFTP LDHTGKPHFK
     101 RDNVYLPGKL LGALNEAAVQ ANVSADTQFT LFLTQDECNP FHDKKRG*

```

The cp6654 nucleotide sequence <SEQ ID 316> is:

```

10     1 ATGAAAACTA AAATGAACTC TAGAAAAAAA GCAGGTCAAT GGGCAATT
      51 CAATTCTCCCA ACTCCTGGTG TCAGTTCAAC TTTAGTTTA GCATGGACTC
     101 CTTGGGGTTA TTACGACAAG GATGTACAAG ATATCTTAGA AAGAAAAGAT
     151 CCGATGAGCT CTTCGCTTC TGAAAAAGAC TCAAAGGAGT TCTTGAAAAAA
     201 TCTGTTTGTA GATCTCTTAG AAAATGGCTT CACATCAGTA CATATTCACG
     251 CAGAAGAACGC TTTCACTCCT CTTGATCATA CGGGAAACC TCACTTTAAA
     301 AGAGACAATG TGTACTTACCG CGGAAAGTTG TTAGGCGCTT TGAATGAGGC
     351 TGCAGGTACAA GCCAATGTAA GTGCGGATAC TCAATTACA TTGTTCCCTA
     401 CTCAGAGATGA GTGCAATCCT TTTCATGATA AGAAAAGAGG TTAA

```

The PSORT algorithm predicts cytoplasm (0.0730).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 154A; 6412 = lanes 2-3; 6431 = lanes 11-12; 6443 = lanes 5-6; 6496 = lanes 8-9; 6654 = lane 10; markers in lanes 1, 4, 7). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 154B, 155, 156, 157 & 158) and for FACS analysis.

These experiments show that cp6412, cp6431, cp6443, cp6496 & cp6654 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from their sequences alone.

Example 159 and Example 160

The following *C.pneumoniae* protein (PID 4376477) was expressed <SEQ ID 317; cp6477>:

```

25     1 LLKFFLVCEE LCILTVATHR ALLETPLALS FFKELKTKYV YRAKDILQLH
      51 NYKGFTILNT SPLCS*

```

The cp6477 nucleotide sequence <SEQ ID 318> is:

```

30     1 TTGCTAAAGT TCTTTCTAGT ATGTGAAGAG TTATGTATAC TTACTGTTGC
      51 TACACATAGA GCTCTCTTAG AAAACTCCTTT AGCTCTATCA TTTTTTAAAG
     101 AACCTAAGAC AAAATATGTC TACAGGGCGA AAGACATACT ACAACTACAT
     151 AACTATAAAAG GATTACTAT CCTTAATACA TCACCGTTAT GTTCTTAA

```

The PSORT algorithm predicts inner membrane (0.128).

35 The following *C.pneumoniae* protein (PID 4376435) was also expressed <SEQ ID 319; cp6435>:

```

1      1 LWSHFPRGFF MLPFCPTILL AKPFLNSENY GLERLAATVD SYFDLGQSQI
      51 VFLSKQDQGI TVEELSAKDR KFKPGSMNCT LYTEDPILPA HNSFSNCSDI
     101 QMRTPISPIH *

```

The cp6435 nucleotide sequence <SEQ ID 320> is:

```

40     1 TTGTGGTCGG ATTCCCCAAG AGGATTTTT ATGCTCCCTT TTTGCCCTAC
      51 CATCCTCTT GCTAAACCTT TTTAAATAG CGAGAAATTAC GGCTTAGAAC
     101 GTTTAGCTGC AACCGTAGAT TCTTATTTG ATCTGGGACA GTCTCAAATA
     151 GTCTTCCTAA GCAAACAGGA TCAAGGAATC ACTGTGGAAG AATTGAGTGC
     201 TAAAGATAGG AAATTCAAGC CAGGCTCTAT GAACTGTACA CTGTACACTG
     251 AAGATCCTAT CTTACCTGCT CATAATTCTT TTAGTAATTG CTCTGATATT
     301 CAAATGCGTA CTCCGATTAG CCCTATACAT TAA

```

-171-

5
1 ATGAGCAGTT CGGAAGTTGT TTTCCAGACA GTTCATGGCC TTGGCTTGG
51 TGATTGTCT TCAAAAAGTG TTGTCCCTTT TAAGAAAAGT CTTTCGGATG
101 CGCCCCGTGT TGTGTGCTCG ATTITAGTTTG TGACTCTGGG GTTGGGAGCG
151 CTTGTTGTG GTATTGCAC TACTTGTGTTGG TGACTCTGGG GAGTTATTTT
201 AATGGGGGGA ATTTGCCTA TAGTTTAGG TGCAATTCTT TTAGCTTTAA
251 GTCTATTCTTG GTTGTGGGGT TTATTTCTA ATTGTGTTGG TTCTAAAGAGA
301 GTTTTACCGG GTGAGGGATT GCTACGGGAT AAGCTTTAG ATGGTGGATT
351 TTCAAGAGCG GCACCTTCAG GAATGGGACT TCCGGGTGAT GGATCTCCAA
401 GAGCGTCAAC GCCATCTGCTAGAGAAC TTCAAGCAGA GATACAGGCA
451 GTTACTCAAG CTATCGATCA GATGTCAGAT GATTGA

The PSORT algorithm predicts inner membrane (0.4864).

The following *C.pneumoniae* protein (PID 4376431) was also expressed <SEQ ID 309; cp6431>:

15
1 LRAGGSLVTT YPKEGQRLRS PEQLRVLDDL VQSYPNHLHA IEILDCGAIPQ
51 DLIGATYIIT FADFSTYIILS LRSYQANSPS DDTWGIWFGS IDDPVQAVIS
101 FLKDHFGLP STLAQDPLLC TNK*

The cp6431 nucleotide sequence <SEQ ID 310> is:

20
1 TTGCGAGCAG GAGGTAGTCT TGTTACAACA TACCCTAAGG AAGGTAGAG
51 ATTGCGCTCC CCAGAACAGT TAAGAGTTCT GGATGATTTA GTGCAAAGCT
101 ATCCAATCA CCTACATGCG ATTGAACCTTG ATTGTGGTGC AATCCCTCAA
151 GATTGATCG GAGCCACCTA TATCATCACG TTCGCGATT TTTCCACCTA
201 TATTCTCTCT TTAAGAAGCT ACCAAGCCAA TTCTCCCTCC GATGATACAT
251 GGGGGATTTG GTTTGGATCT ATTGACGATC CTGTTCAAGC AGTCATATCA
301 TTTTAAAAG ATCATGGATT TGCTCTTCCC TCGACCTTAG CTCAAGATCC
351 TTGCTTTGT ACTAACAAAGT AA

25 The PSORT algorithm predicts cytoplasm (0.2115).

The following *C.pneumoniae* protein (PID 4376443) was also expressed <SEQ ID 311; cp6443>:

30
1 MIMTTISNSP SPALNPESL IPPPTLVSSG TQTSLAYTIP AQGRRSTLRI
51 ILDIFIIILG LATIISTFIV IFFLNGLNLL STPSIISSSC LIIVGLLFLI
101 MGLYFMISSL DQGLVGLLQK ELSQAEEREE EVIQEIEALR GAPRAESPTE
151 SPSTWL*

The cp6443 nucleotide sequence <SEQ ID 312> is:

35
1 ATGATTATGA CTACTATATC TAACTCACCC TCCCCTGCAT TGAATCCGA
51 ACTTTCCCTT ATTCCTCCAC CAACACTTGT ATCTTCAGGT ACGCAACAT
101 CTCTAGCTTA TACGATCCCC GCACAAGGAC GAAGATCCAC CCTACGTATT
151 ATATTAGATA TATTCTATT CATTCTTGGT TTAGCTACGA TCATTCTAC
201 CTTTATTGTT ATTTCTTT TAAATGGGCT GAACCTTGCTC TCGACCCCCAT
251 CTATTATCTC TTCGTCATGT TTAATCATTG TTGGATTGCT TTTTTTGATT
301 ATGGGGTTAT ATTCATGAT CTCGAGTTTG GATCAGGGGC TTGTAGGCCT
351 TCTGAAAAG GAACTCTCTC AAGCCGAAGA AAGAGAAGAA GAGTATATCC
401 AGGAAATCGA AGCTTTAAGA GGAGCTCTA GAGCAGAATC TCCCACAGAG
451 TCTCCTAGTA CCTGGTTATG A

The PSORT algorithm predicts inner membrane (0.5585).

The following *C.pneumoniae* protein (PID 4376496) was also expressed <SEQ ID 313; cp6496>:

45
1 MLIGRYSSDD QFTEATKNTP TIKLGFRD NLEGLTNPIS EIVSETSSSI
51 KDSVLRSLPI LGSILGCARL YSTLSTNDPL DETQEKIWHT IFGALETLGL
101 GILILLFKII FVILHCFHL VIGFCK*

The cp6496 nucleotide sequence <SEQ ID 314> is:

50
1 ATGCTAATAG GCAGATACAG TAGTGATGAC CAATTCACTG AAGCAACAAA
51 AACACACCCA ACCATAATTA AGCTAGGTT TGTTAGAGAT AATCTCGAGG
101 GATTAACGAA CCCTATCTCT GAAATCGTCT CGGAAACCTC CTCTTCTATT
151 AAAGATTCCG TTCTTCGCTC TCTTCTATT TTAGGGTCCA TTTTAGGATG
201 CGCCCGACTT TACAGCACAC TCTCTACAAA TGATCCTCTT GACGAAACTC
251 AAGAAAAGAT TTGGCACACT ATATTGAGG CCTTAGAAC CTTAGGCTTA
301 GGGATTCTCA TCCTCTTATT TAAAAATTATT TTTGTTATAT TACACTGCAT
351 ATTCATCTA GTTATTGGGT TCTGCAAATA A

-174-

5 1 MRPHRKHVSS KSLALKQSAS THVEITTKAF RLSMPLKQLI LEKSDHLPPM
 51 ETIRVVLTSK KDKLGTEVHV VASHGKEILQ TKVHNANPYT AVINAFKKIR
 101 TMANKHSNKR KDRTKHDLGL AAKERIAIQ EEQEDRLSNE WLPVEGLDAW
 151 DSLKTLGYVP ASAKKKISKK KMSIRMLSQD EAIRQLESAA ENFLIFLNEQ
 201 EHKGQCIYKK HDGNVVLIEP SLKPGFCI*

The cp6881 nucleotide sequence <SEQ ID 326> is:

10 1 ATGAGACCTC ATCGTAAACA CGTATCATCT AAAAGCTTAG CTTTAAAGCA
 51 ATCTGCATCA ACTCATGTAG AGATCACAAC AAAAGCCTTT CGTCTCTCTA
 101 TCCCTCTAAA ACAGCTGATC CTAGAGAAAA GCGACCACCT CCCCCCTATG
 151 GAAACAATCC GTGTGGTGCT AACCTCTCAT AAAGATAAGC TAGGCACCGA
 201 GGTGCATGTT GTAGCTTCAT ATGGCAAAGA AATCCTTCAA ACTAAGGTTC
 251 ATAACGCAAA CCCATACACT GCAGTGATCA ATGCTTTAA GAAAATCCGC
 301 ACCATGGCAA ATAAGCACTC CAATAAACGT AAAGACAGGA CAAAACATGA
 351 TCTAGGTCTT GCAGCAAAAG AAGAACGTAT CGCAATACAG GAAGAACAAAG
 401 AAGATCGCCT TAGCAACGAG TGGCTTCCTG TCGAAGGCCT CGATGCCCTGG
 451 GATTCTCTAA AAACCTTTGG GTATGTTCCC GCATCAGCGA AAAAGAAGAT
 501 CTCCAAGAAA AAGATGAGCA TTCTGTATGCT ATCTCAAGAC GAGGCTATCC
 551 GCCAGCTAGA GTCTGCCGCA GAAAACTTCC TGATCTTCTT GAACGAGCAA
 601 GAGCATAAAA TCCAATGCAT TTATAAAAAA CATGACGGCA ACTATGTCCT
 651 TATTGAACCT TCCCTCAAGC CAGGATTCTG CATCTGA

The PSORT algorithm predicts cytoplasm (0.249).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 161A; 6441= lanes 7-9; 6748 = lanes 2-3; 6881 = lanes 4-6). The recombinant protein was used to immunise mice, whose sera were used in Western blots (Figures 161B, 162 & 163) and for FACS analysis.

25 These experiments show that cp6441, cp6748 & cp6881 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequence alone.

Example 164 and

Example 165

Example 166

The following *C.pneumoniae* protein (PID 4376444) was expressed <SEQ ID 327; cp6444>:

1 MEQPNCVIQD TTTVLYALNS FDPRLSDDTH RLGKQSPLEA ENALGEFIEG
 51 LDTNSFPLEE VAIPILPGYH PKFYLSFIDR DDQGVHYEVN DGVFLKTVAA
 101 CIIENSFLTD SMSPELLSEV KEALKR*

35 The cp6444 nucleotide sequence <SEQ ID 328> is:

40 1 ATGGAGCAAC CCAATTGTGT GATTCAGGAT ACTACAAC TG TTTGTATGC
 51 CTTAAATAGC TTTGATCCTA GACTTAGTGA TGACACTCAC AGACTTGGGA
 101 AGCAATCACC TCTTGAGCA GAAAATGCTC TTGGAGAATT TATTGAAGGT
 151 TTGGATACAA ATAGCTTCC TTTAGAGGAA GTTGCCATTC CCATCCTGCC
 201 AGGTTATCAC CCTAAGTTTT ATTATATCTT CATAGATAGG GACGATCAAG
 251 GTGTCCACTA TGAAGTTTA GATGGCGTAT TTTTAAAGAC AGTCGCTGCT
 301 TGTATTATAG AGAACTCCCTT CTTAACTGAT TCTATGAGCC CGGAGCTTCT
 351 CAGCGAAGTT AAGGAAGCTC TGAAACGATG A

The PSORT algorithm predicts cytoplasm (0.2031).

45 The following *C.pneumoniae* protein (PID 4376413) was also expressed <SEQ ID 329; cp6413>:

1 MAVQSIKEAV TSAATSVGCV NCSREAIPIAF NTEERATSIA RSVIAIIAV
 51 VAISLLGLGL VVLAGCCPLG MAAGAITMLL GVALLAWAIL ITLRLNNIPK
 101 AEIPSPGNNG EPNERNNSATP PLEGGVAGEA GRGGGSPLTQ LDLNSGAGS*

The cp6413 nucleotide sequence <SEQ ID 330> is:

50 1 ATGGCTGTTA AATCTATAAA AGAAGCCGTA ACATCAGCCG CAACATCAGT

The PSORT algorithm predicts periplasmic space (0.4044).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 159A; 6435 = lanes 2-4; 6477 = lanes 5-7). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 159B & 160) and for FACS analysis.

- 5 These experiments show that cp6477 & cp6435 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequences alone.

Example 161 and

Example 162 and

Example 163

- 10 The following *C.pneumoniae* protein (PID 4376441) was expressed <SEQ ID 321; cp6441>:

```

1  VEAGANVLVI DTAHAHSKGV FQTGLEIKSQ FPQISLVVGN LVTAAAVSL
51 AEIGVDAVKV GIGPGSICTT RIVSGVGYPQ ITAITNVAKA LKNSAVTIA
101 DGRIRYSGDV VKALAAGADC VMLGSLLAGT DEAPGDIVSI DEKLFKRYRG
151 MGSLGAMKQG SADRYFQTQG QKKLVPGGVE GLVAYKGSVH DVLYQILGGI
201 RSGMGYVGAE TLKDLTKAS FVRITESGRA ESHIHNIYKV QPTLNY

```

The cp6441 nucleotide sequence <SEQ ID 322> is:

```

1  GTGGAAGCTG GAGCAAATGT TCTAGTCATT GACACAGCTC ATGCACACTC
51 TAAAGGAGTA TTCCAAACAG TTTTAGAAAT AAAATCCCAG TTCCCACAAA
101 TTTCTTTAGT TGTAGGGAAAT CTTGTTACAG CTGAAGCCGC AGTTTCCTTA
151 GCTGAGATTG GAGTTGACGC TGTAAGGTA GGTATTGGCC CAGGATCTAT
201 CTGTACAACCT AGAACATCGTT CAGGGGTCGG TTATCCACAA ATTACTGCCA
251 TTACAAACGT AGCAAAAGCT CTTAAAAACT CTGCCGTGAC TGTAATTGCT
301 GATGGGAGAA TCCGCTATTG TGGAGATGTG GTAAAAGCAT TAGCAGCAGG
351 ACCAGACTGT GTCATGCTAG GAAGTTTGCT TGCAGGGACT GATGAAGCTC
401 CTGGGGATAT CGTTTCTATC GATGAGAAC TTTTTAAAAG GTACCGCGGC
451 ATGGGATCTT TAGGCGCTAT GAAACAAGGA AGTGCTGACC GGTATTTCA
501 AACACAGGGGA CAGAAAAGC TGGTTCCCTGG GGGAGTTGAA GGACTAGTCG
551 CTTATAAAAGG CTCTGTCCAC GATGTCCTCT ATCAAATTTC AGGAGGAATA
601 CGCTCAGGTA TGGGGTATGT TGGAGCTGAA ACTCTCAAAG ATTTAAAAAC
651 TAAGGCTTCC TTTGTTGCAA TTACTGAATC TGGAAAGAGCT GAAAGTCATA
701 TTCATAATAT TTACAAAGTT CAACCAACCT TAAATTATTA A

```

The PSORT algorithm predicts bacterial inner membrane (0.132).

- The following *C.pneumoniae* protein (PID 4376748) was also expressed <SEQ ID 323; cp6748>:

```

1  LFSEGTLALNL FRIFAPLNRN VTTEYSRARQ PDLHRIAIVY IGVLDSESSK
51 ILERLISYMS CIYSESQMYL RFFMGKVNQ SAVLSKLHVE NLHIRCGFFS
101 EDAVPESEPF DLSIYVHTDR SCPLPTKKRS SSWELOQTVEL PESIYPQSEF
151 LLMRPRMLS*

```

The cp6748 nucleotide sequence <SEQ ID 324> is:

```

1  TTGTTCTCTG AGGGGACAGC TCTAAATTAA TTTCGTATAT TTGCTCCACT
51 ACGCAACCGT GTGACTACAG AATACAGTCG TGCTAGGCAA CCCGACCTAC
101 ATAGAATTGC CATCGTCTAT ATAGGAGTTC TCGATTCAAG AAGTTCCAAG
151 ATCCCTAGAGC GGCTAATCTC TTATATGAGT TGTATCTATT CTGAATGCCA
201 AATGTATTTA AGATTCTTAA TGGGAAGAA TGTAATCAA AGTGCCTGTAC
251 TCTCAAAATT ACATGTAGAA AATCTGCACA TCCGTTGTGG GTTTTCAGC
301 GAGGATGCTG TTCCAGAGAC TGAGCCCTTC GATCTCTCCA TCTACGTGCA
351 CACAGATCGT AGCTGTCTC TCCCTACGAA AAAACGGAGC AGCTCCTGGG
401 AACTCCAAAC TGTAGAACTC CCAGAGTCAA TATATCCACA GTCGGAATTG
451 CTATTGATGA GACCTCGAAT GCTTTCGTAG

```

The PSORT algorithm predicts cytoplasm (0.170).

- 50 The following *C.pneumoniae* protein (PID 4376881) was also expressed <SEQ ID 325; cp6881>:

5
201 AACAGAGAAG ACCACGACCC GTCATTGGT GCTCTCTATT CGCCATAACG
251 CCTCTCTTAT TGTAATTCTGT ACGGTTCCCTG GTTCAGCTTC TTGGATCGCT
301 GCTTGTGTTAG ATCAAGGGCT CAAAGATGAA ATTCTTGAACT CTTTGGCAGG
351 AGATGACACG ATTTTTGTCA CTCCATAGA TGAAGGGAGG CTCCCATTGT
401 TGATGGTTTC GATTGCAAAT TTACTGCAAG TTTTCTTGAA TTAA

The PSORT algorithm predicts inner membrane (0.1510).

The following *C.pneumoniae* protein (PID 4376540) was also expressed <SEQ ID 335; cp6540>:

10
1 MSQCQSSSTS TWEWMKSFVP NWKNPTPLS PIPSEDEFIL AYEPFVLPKT
51 DPENAQANPP GTSTPNVENG IDDLNPLLQ PNEQNNANNP GTSGSNPTSL
101 PAPERLPETE ENSQEEEQGS QNNEDLIG*

The cp6540 nucleotide sequence <SEQ ID 336> is:

15
1 ATGTCTCAAT GTCAGAGTAG CAGTACATCT ACCTGGGAAT GGATGAAATC
51 TTTTGTGCCA AACTGGAAGA ATCCAACCTCC CCCCTTATCT CCTATACCTT
101 CTGAGGACGA ATTTATATTA GCATACGAGC CATTGTTCT ACCGAAAACA
151 GATCCAGAAA ACGCACAAAGC TAATCCTCCA GGCACATCTA CACCGAATGT
201 AGAAAACGGG ATCGATGATC TCAACCCCTCT TCTGGGGCAA CCCAACGAAC
251 AAAACAATGC CAACAATCCA GGAACCTCTG GATCTAATCC TACATCTCTA
301 CCCGCCCCCG AACGACTCCC TGAAACTGAA GAGAACAGCC AAGAAGAAGA
351 ACAAGGATCT CAAAATAATG AGGATCTTAT AGGATAA

20 The PSORT algorithm predicts cytoplasm (0.3086).

The following *C.pneumoniae* protein (PID 4376743) was also expressed <SEQ ID 337; cp6743>:

1 LREEGSVSFR EYFRAYMCDK IVAQKNFLFT LDRAVIKQAGW RSQEKLNLFY
51 VESQALGREI KVSLEYYIQS MVGILGSQRT KKSFKFSVDF TPLEQALQER
101 CSSDDDEDAT ATSTATGATA SPTDMHEDE*

25 The cp6743 nucleotide sequence <SEQ ID 338> is:

30
1 TTGAGAGAAG AAGGTAGTGT TTCTTTCAAGA GAATATTTCA GAGCCTATAT
51 GTGTGATAAA ATCGTGGCAC AGAAGAACTT CTTATTTACT TTAGACGCTG
101 TAATTAAACAA GGCCGGTTGG AGATCACAAG AGAAACTCAA TTTATTTAT
151 GTTGAAAGTC AGGCTTTAGG AAGAGAAATC AAAGTCAGCT TAGAGGAATA
201 TATTTCAGAGT ATGGTCGGGA TTTTGGGATC TCAGAGAACCC AAGAAAAGCT
251 TTAAGTTTTTC TGTCGACTTT ACCCCCTTTAG AGCAGGCTCT ACAAGAAAGA
301 TGCTCTCTG ATGATGACGA AGATGCAACA GCAACTTCGA CCGCTACAGG
351 GGCAACAGCA TCTCCGACTG ACATGCACGA AGATGAGTAA

The PSORT algorithm predicts cytoplasm (0.2769).

35 The following *C.pneumoniae* protein (PID 4377041) was also expressed <SEQ ID 339; cp7041>:

40
1 MLMMLMMIIG ITGGSGAGKT TLTQNIKEIF GEDVSVICQD NYYKDRSHYT
51 PEERANLIWD HPDAFDNDLL ISDIKRLKNN EIVQAPVPDF VLGNRSKTEI
101 ETIYPSKVIL VEGILVFENQ ELRDLMDIRI FVDTDADDERI LRRMYRDVQE
151 QGDSVDCIMS RYLSMVKPMH EKFIIEPTRKY ADIIVHGNYR QNVVTNILSQ
201 KIKNHLENAL ESDETYYMVN SK*

The cp7041 nucleotide sequence <SEQ ID 340> is:

45
1 ATGTTGATGA TGCTTATGAT GATTATTGGA ATTACAGGAG GTTCTGGAGC
51 TGGGAAAACC ACCCTAACCC AAAACATTAA AGAAATTTC GGTGAGGATG
101 TGAGTGTAT CTGCCAAGAT AATTATTACA AAGATAGATC TCATTATACT
151 CCTGAAGAAC GTGCCAATT TATTGGGAT CATCCGGACG CCTTTGATAA
201 TGACTTTATTA ATTTCAAGACA TAAAACGTCT AAAAATAAT GAGATTGTCC
251 AAGCCCCAGT TTTTGATTTT GTTTAGGTA ATCGATCTAA AACGGAGATA
301 GAAACGATCT ATCCATCTAA AGTTATTCTT GTTGAAGGTA TTCTGGTCTT
351 TGAAGATCAA GAACCTAGAG ATCTTATGGA TATTAGGATC TTTGTAGACA
401 CCGATGCTGA TGAAAGGATA CTACGCCGTA TGGTTCGAGA TGTTCAAGAA
451 CAAGGAGATA CGCTGGACTG CATCATGTCT CGTTATCTT CTATGGTAAA
501 GCCTATGCAT GAGAAATTAA TAGAGCCGAC TCGGAAATAT GCTGATATCA
551 TTGTACATGG AAATTACCGA CAAAACGTAG TAACAAATAT TTTGTCACAG
601 AAAATTAAAA ATCATTAGA GAATGCCCTG GAAAGCGATG AGACGTATTA
651 TATGGTCAAC TCTAAGTAA

5 51 AGGATGTGTA AACTGTTCTA GAGAGGCTAT ACCAGCATT AATACAGAGG
 101 AGAGAGCAAC GAGTATTGCT AGATCTGTTA TAGCAGCTAT CATTGCTGTT
 151 GTAGCTATCT CCTTACTCGG ACTAGGTCTT GTAGTTCTTG CTGGTTGCTG
 201 TCCTTTAGGA ATGGCTGCGG GTGCATAAAC AATGCTGCTG GGTGTAGCAT
 251 TATTAGCTTG GGCAATACTG ATTACTTTGA GACTGCTAA TATAACCTAAG
 301 GCTGAAATAC CGAGTCAGG GAACAACGGT GAGCCTAATG AAAGAAATTC
 351 ACGAACTCCT CCTCTAGAGG GTGGTGTTGC AGGAGAACGCC GGTGCGGGCG
 401 GGGGGTCACC TTTAACCCAA CTTGATCTCA ATTCAAGGGGC GGGAAAGTTAG

The PSORT algorithm predicts inner membrane (0.6180).

- 10 The following *C.pneumoniae* protein (PID 4377391) was also expressed <SEQ ID 331; cp7391>:

1 1 MMLRVIELPL LPIKQALEKA FVQYNSYKAK LTKVEPCFRE SPAYITSEER
 51 LQSLDQTLER AYKEYQKRFQ EPSRLESEVS GCREHLREQV KQFETQQLDL
 101 IKEELIFVSD VLFRKVMVSCL VSTVHVPFME FYYEYFELHR LRLRAQWMAN
 151 AEIYSKVRKA FPEMLKETLE KAKAPREEEY WLCEERKS K EKRLILNKIE
 201 AAQQRVKDLE PPPIKETGKQ KRKKEYSFFI RLKS*

The cp7391 nucleotide sequence <SEQ ID 332> is:

20 1 ATGATGCTTC GTGTACATAGA GCTTCCACTA CTTCCTATAA AGCAAGCGTT
 51 GGAGAAGGCT TTTGTACAAT ATAATAGCTA CAAAGCGAAG TTAACCAAGG
 101 TAGAACCTTG CTTTAGAGAG AGCCCTGCCT ATATAACTAG CGAACAGCGA
 151 CTCCAGAGTT TGGATCAGAC TTTAGAACGT GCGTACAAAG AGTACCCAGAA
 201 GAGATTCCAG GAGCCTTCAC GTTTGGAATC GGAAGTAAGT GGATGTAGAG
 251 AGCATCTTAG AGAGCAGGTA AAACAATTG AAACCTCAAGG ACTAGACTTG
 301 ATCAAAGAAG AGCTTATTTT TGTTAGTGAT GTGTTATTCC GAAAATGGT
 351 CAGTTGCTCA GTGTCGACAG TGCATGTTCC CTTTATGGAG TTTTATTATG
 401 AGTATTTGAG TTGTCATAGA TTGAGGTTGC GGGCCCAATG GATGGCGAAT
 451 GCGGAGATT ATAGCAAAGT TAGAAAAGCA TTCCCAGAGA TGTGGAAGGA
 501 GACCTTAGAA AAAGCTAAGG CTCCCAGAGA AGAAGAGTAT TGGTTACTTT
 551 GCGAGGAGAG AAAGAGTAAG GAGAAGCGTT TGATTCTCAA CAAGATAGAG
 601 GCAGCTCAGC AGCGGGTAAA AGATTTAGAA CCTCCTCTA TAAAGAGAC
 651 AGGGAAACAG AAACGGAAGA AAGAATATTC GTTTTCATT CGATTAAT
 701 CGTGA

The PSORT algorithm predicts inner membrane (0.1489).

The proteins were expressed in *E.coli* and purified as his-tag and GST-fusion products (Figure 164A; 6444=lanes 11-12; 7391=lanes 2-3; 6413=lanes 4-6). The recombinant protein was used to immunise 35 mice, whose sera were used in Western blots (Figures 164B, 165 & 166) and for FACS analysis.

These experiments show that cp6444, cp6413 & cp7391 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequence alone.

40 **Example 167**,
 Example 168,
 Example 169 and
 Example 170

The following *C.pneumoniae* protein (PID 4376463) was expressed <SEQ ID 333; cp6463>:

45 1 MKKVTIDEA LKEILRLEGA ATQEELCAKL LAQGFATTQS SVSRWLRKIQ
 51 AVKVAGERGA RYSLPSSTEK TTTRHLVLSI RHNASLIVIR TVPGSASWIA
 101 ALLDQGLKDE ILGTLAGDDT IFVTPIDEGR LPLLMVSIAN LLQVFID*

The cp6463 nucleotide sequence <SEQ ID 334> is:

50 1 ATGAAAAAAA AAGTAACATAT AGATGAGGCT TTAAAAGAAA TTTTACGTCT
 51 TGAAGGAGCG GCAACTCAGG AGGAATTATG TGCAAAACTC TTAGCTCAAG
 101 GTTTTGCTAC AACCCAGTCG TCTGTATCTC GTTGGCTACG AAAGATTTCAG
 151 GCTGTAAAGG TTGCTGGAGA CGGTGGTGCT CGTTATTCTT TACCTCTTC

201 TTGGTATACA AGTGACCAAG ATTGGAAAAA ACAAGTGGTT TGA

The PSORT algorithm predicts inner membrane (0.145).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 171A; 6632 = lanes 5-7; 6648 = lanes 8-10; 6497 = lanes 2-4). The recombinant proteins were used to immunise mice,

5 whose sera were used in Western blots (Figures 171B, 172, 173) and for FACS analysis.

These experiments show that cp6632, cp6648 and cp6497 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequence alone.

Example 174,

Example 175,

Example 176,

Example 177 and

Example 178

The following *C.pneumoniae* protein (PID 4377200) was expressed <SEQ ID 347; cp7200>:

15 1 MPVPIDNSSL NLQEVPESLE DLEQHAEESP THQSAESSSL QLSLASSAIS
 51 SRVEQLSSLV LGMENSDFSS LRDVPIFSAI YESSTHTPVP TPLVGVGYIN
 101 GSQSGYYDTQ RESLHLSQLL GSRRVEVVYN QGNFMEASLL NLCPRRPRRD
 151 PSPISLALLE LWEAFFLEHP PGSTFNPIFF W*

The cp7200 nucleotide sequence <SEQ ID 348> is:

20 1 ATGCCCGTTC CTATAGATAA TTCCCTCTCGC AACCTACAAG AAGTTCCAGA
 51 AAGCCTAGAA GACCTCGAAC AACACGCAGA AGAACATCCT ACTCATCAAA
 101 GTGCAGAAAG CAGTTCTTTG CAACTGTCTC TAGCCTCCTC AGCAATTTCAGA
 151 AGTAGAGTAG AACAACTATC TTCCCTCGTC TTAGGAATGG AAAATTTCAGA
 201 TTTCTCCCTCT TTAAGAGACG TTCCCTATCTT CTCAGCTATC TACGAATCTT
 25 251 CAACACACAC ACCTGTCCCC ACTCCTCTAG TTGGCGTGGG ATATATCAAC
 301 GGAAGTCAAT CAGGATACTA CGATACACAA AGAGAACATCTC TTCACCTCAG
 351 CCAATTGTTA GGAAGCGAA GAGITGAAGT TGTCTATAAC CAAGGAAACT
 401 TCATGGAGGC CTCTTTGCTA AATCTGTGCC CCAGAAGACC TCGAAGAGAT
 45 451 CCCCTCTCCAA TTTCTTTAGC TCTATTAGAG CTCTGGGAAG CATTTTTTTT
 501 AGAACACCCCC CCAGGTAGCA CTTTTAATCC AATATTTTTT TGGTAA

The PSORT algorithm predicts cytoplasm (0.3672).

The following *C.pneumoniae* protein (PID 4377235) was also expressed <SEQ ID 349; cp7235>:

35 1 LNFVSTLTGS DFYAPVLEKL EEAFADETTGQ VILFSSSPDF IVHPIAQQLG
 51 ISSWYASCYR DQSAEQTIYK KCLTGDKKAQ ILSYIKKINQ ARSHTFSDHI
 101 LDLPFLMLGE EKTVVVRPQGR LKKMAKKYYW NIV*

The cp7235 nucleotide sequence <SEQ ID 350> is:

40 1 TTGAATTTTG TATCGACTCT GACCGGCTCC GATTTTTATG CTCCTGTTTT
 51 AGAAAAAACTA GAAGAACGCTT TTGCAGATAC CACAGGACAG GTGATCCTTT
 101 TTTCTTCITTC TCCAGACTTT ATTGTCCACC CCATAGCGCA GCAACTCGGG
 151 ATTAGTTCTT GGTATGCGTC GTGTTATCGC GATCAGTCTG CAGAACAGAC
 201 GATCTATAAA AAATGTCCTA CAGGGGATAA AAAAGCGCAA ATTTTGAGTT
 251 ATATTAACAA AATTAATCAA GCAAGAAGCC ATACCTCTC CGACCATATT
 301 TTAGATCTTC CTTTTCTTAT GCTGGGGAGAA GAGAAAACCG TCGTTCGCCC
 351 TCAGGGACGA CTCAAGAAAA TGGCAAAAAA ATATTACTGG AATATCGTTT
 401 AA

The PSORT algorithm predicts cytoplasm (0.3214).

The following *C.pneumoniae* protein (PID 4377268) was also expressed <SEQ ID 351; cp7268>:

1 MMHRYFIPLL ALLIFSPSLV RAELOPSEN R KGGWPTQLSC AEGSQLFCKF

The PSORT algorithm predicts inner membrane (0.1022).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 167A; 6463 = lanes 2-4; 6540 = lanes 5-7; 6743 = lanes 8-9; 7041 = lanes 10-11). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 167B, 168, 169 & 170) and for FACS analysis.

These experiments show that cp6463, cp6540, cp6743 & cp7041 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequence alone.

**Example 171 and
Example 172 and
Example 173**

The following *C.pneumoniae* protein (PID 4376632) was expressed <SEQ ID 341; cp6632>:

```

1 VQLFQYMNES GWDWLCDFDS QGEGFQLSRL VGLLHSSWAL YEAKEQFYLP
51 EVSLLTWEEL IEMQLLSKPT KHGVAKDLCN VFEKFHFQRFR QYLGSLDLNQ
101 RFENTFLNYP KYHLDRE*

```

The cp6632 nucleotide sequence <SEQ ID 342> is:

```

1 GTGCAATTAT TTCAATATAT GAATGAGTC GGATGGGATT GGCTTTGTGA
51 TTTTGATTCT CAAGGGCAGG GATTCCAGTT ATCACGTC TGTTGGCTGT
101 TACATTCGTC CTGGGCATTA TACGAAGCAA AAGAGCAATT TTACCTTCCT
151 GAGGTTTCTC TATTGACCTG GGAAGAACTG ATAGAAATGC AGTTATTAAAG
201 CAAACCAACA AAACACGGGG TTGCAAAAGA TCTTTGTAAT GTATTGAAA
251 AACACTTCA AAGGTTPTAGA CAGTACCTAG GTTCCTTAGA TCTAAATCAA
301 AGGTTCGAAA ATACCTTCTT GAATTATCCT AAATACCATT TAGATAGGGA
351 GTGA

```

The PSORT algorithm predicts cytoplasm (0.3627).

The following *C.pneumoniae* protein (PID 4376648) was also expressed <SEQ ID 343; cp6648>:

```

1 MPVSSAPLPT SHRPSSGNLG LMEPNSKALK AKHQDKTTKT IKLLVKILVA
51 ILVIEVLGII AAFFFPGTPP ICLLILGGLI LTTVLCVLLL VIKLALVNKT
101 EGTAAEQQIK RKLSSKSIS*

```

The cp6648 nucleotide sequence <SEQ ID 344> is:

```

1 ATGCCCGTGT CCTCAGCCCC CCTACCCACA AGCCACCGCC CTTCCCTCTGG
51 AAATCTAGGC CTCATGGAAC CAAATTCCAA AGCTCTAAAA GCAAAGCATC
101 AAGATAAAAC GACGAAGACG ATTAAACTTT TAGTTAAAAT CCTTGTTGCC
151 ATTCTAGTAA TAGAAGTTTT AGGAATAATT GCAGCTTCT TTATTCCTGG
201 GACTCCTCCC ATCTGCTTGA TTATCCTAGG AGGCCTTATT CTTACAACAG
251 TACTCTGTGT GCTTCTTCTT GTTATAAACG TTGCCCTTGT AAACAAAACC
301 GAAGGAACAA CTGCTGAACA GCAGATAAAA CGTAAACTCT CTTCTAAAAG
351 TATTCTTAG

```

The PSORT algorithm predicts inner membrane (0.6074).

The following *C.pneumoniae* protein (PID 4376497) was also expressed <SEQ ID 345; cp6497>:

```

1 MKPNSTIIFLE NTKHYPDIFR EGFVRDRHGL MEASDWLLST EITIIRSILG
51 AIPILGNILG AGRLYSVWYT SDEDWKKQVV *

```

The cp6497 nucleotide sequence <SEQ ID 346> is:

```

1 ATGAAGCCAA ATAGTATTAT TTTTTTAGAA AATACTAACG ATTATCCCGA
51 CATCTTCGA GAAGGATTG TTCGTGATCG TCATGGACTA ATGGAAGCCT
101 CGGATTGGTT ACTTTCTACG GAAATTACGA TCATTCGCTC CATTCTGGGA
151 GCTATCCCTA TTTTAGGAAA TATTCTTGGGA GCCGGACGAC TCTATAGCGT

```

	101	TAGTTTTGG TATGCTCTTA CTGATTCAG GAGCTCTCTT TCTGACGTTA
	151	GGGATTCCAG GATTGAGTGC AGCAATTCT TTTGGATTAG GCATCGGTCT
5	201	CTCCGCATTA GGAGGAGTGC TGATGATTTC GGGACTACTA TGTCTTTAG
	251	TAAAACGAGA GATTCCGACA GTACGACCAG AAGAAATTCC TGAAGGGGTT
	301	TCGCTGGCTC CTTCTGAGGA GCCAGCTCA CAGGCAGCTC AGAAGACTTT
	351	AGCTCAGCTG CCTAAGGAAT TGGATCAGTT AGATACAGAT ATTCAAGGAAG
	401	TGTTCCGATG TTAAAGAAAG CTGAAAGATT CTAAGTATGA AAGTCGAAGT
	451	TTTTTAAACG ATGCTAACAA GGAGCTTCGA GTTTTGACT TTGTGGTTGA
10	501	GGATACCCCTC TCGGAGATTT TCGAGTTGCCG GCAGATTGTG GCTCAAGAGG
	551	GATGGGATT AAACCTTTTG ATCAATGGGG GACGAAGCCT CATGATGACT
	601	GCAGAACTCG AATCGCTGA TTGTTTCAT GTATCGAAGC GGCTAGGGTA
	651	TTTACCTCT GGGGATGTC GAGGGGAGGG GTTAAAGAAA TCTGCGAAGG
	701	AGATAGTCGC TCGTTGATG AGCTTGCATT GCGAGATTCA CAAGGTGGCG
15	751	GTAGCGTTG ATAGGAATT CTTATGCGATG GCAGAAAAGG CGTTTGCAGAA
	801	ACCGTTGGGA GCTTTAGAAG AGAGTGTGTA TCGGAGTCG ACGCAGAGTT
	851	ATAGAGATAA ATTTTGGAG AGCGAGAGGG CGAAGATCCC ATGGAATGGG
	901	CATATAACCT GTTTAAGAGA TGATGCGAAG AGTGGGTGTG CTGAAAAGAA
	951	GCTTCGGGAT GCGGAGGAAC GTTGGAAAGAA ATTTAGGAAA GCAGTCTTTT
20	1001	GGGTAGAAGA AGACGGGGC TTGACATCA ATAATCTCT TGGAGACTGG
	1051	GGGACAGTGC TTGATCCTTA TAGACAAGAG AGAATGGACG AGATAACGTT
	1101	CCATGAGTTG TATGAAAAAA CTACGTTTT GAAAAGACTG CACAGAAAGT
	1151	GTGCGTTAGC GAAAACAACC TTGAAAAGA AGAGATCTAA AAAGAATTG
	1201	CAGGCAGTCG AGGAGGCCA TGACGTAGG TTGAAATATG TAAGGGATTG
25	1251	GTATGATCAG GAGTTTCAGA AAGCAGGGGA GAGATTAGAG AAACCTGCATG
	1301	CTTGTATCC TGAGGTTCA GTCTCTATAA GAGAGAACAA AATACAAGAG
	1351	ACCGCCTCTA ATTTAGAGAA AGCCTATGAG GCTATCGAAG AGAACTATCG
	1401	TTGCTGTGTC CGAGAGCAAG AGGACTACTG GAAAGAAGAA GAGAAAAGGG
	1451	AAGCGGAGTT TAGGGAGAGG GGAAACAAGA TTCTTCTCC TGAGGAGCTG
30	1501	GAAAGTTCTT TGGAGCAATT CGACCATGGT TTGAAAATT TTTCTGAGAA
	1551	ATTAATGGAA TTGGAAGGGC ATATCTAAA ACTTCAGAAA GAAGCCACAG
	1601	CAGAGGTGGA GAATAAAA CTTTCAGATG CAGAGAGCCG CCTTGAGATT
	1651	GTATTGAAAG ATGTCAGGAGA GATGCCCTGT CGAATTGAGG AGATAGAGAA
	1701	GACCGTGCCT ATGGCGGAGC TGCCCCCTACT TCCTACGAAG AAGCGTTTG
35	1751	AGAAGGCCTG CTCACAAATAT AATAGCTGC CAGAGATGTT GGAGAAGGTG
	1801	AAGCCTTACT GCAAGGAGAG CCTCGCCTAT GTGACTAGCA AAGAGCGTTT
	1851	AGTGAGCTTG GATGAAGATT TACGACGAGC CTACACAGAG TGTCAAGAAGA
	1901	GATTCCAGGG GGATTCGGGT TTGGAGTCGG AAGTAAGAGC CTGTCAGAG
	1951	CAACTGCGAG AGCGGATCCA AGAGTTGAA ACTCAAGGGC TGGACTTGGT
40	2001	GGAAAAAGAG TTGTTTGTG TGAGTAGTAG ATTAAGAAAT ACAGAGTGC
	2051	ATTGTGTATC TGGTGTAAAG AAAAGAACAC CCTCTGGTAA GAAAGTTTAT
	2101	GCCCAGTATT ATGATGAGAT TTATCGAGTT AGAGTTCAAT CCCGATGGAT
	2151	GACGATGTCT GAGAGATTGA GAGAGGGAGT TCAAGCATGC AACAAAGATGT
	2201	TGAAGGCAGG CCTAACGGAA GAAGATAAGG TTCTTAAAGA AGAAGAGTAT
45	2251	TGGTTGTACT GAGAGGAGAG AAAAGATAAA GAGAAAACGTT TGGTTGGTAC
	2301	TAAGATAGTA GCAACGCGAGC AGCGAGTTGC AGCATTGAA TCCATAGAAG
	2351	TTCTGAGAT TCCTGAGGCC CCAGAGGAGA AACCGAGTTT GCTGGATAAA
	2401	GGCGGTTCTT TATTTACTCG CGAGGACCAT ACCTAG

The PSORT algorithm predicts inner membrane (0.461).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 174: 7200=lanes 2-3; 50 7236=lanes 4-5; 7268=lanes 6-8; 7375=lanes 9-10; 7388=lanes 11-12). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 174, 175, 176, 177 & 178) and for FACS analysis.

These experiments show that cp7200, cp7235, cp7268, cp7375 & cp7388 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident 55 from the sequence alone.

Example 179

The following *C.pneumoniae* protein (PID 4376723) was expressed <SEQ ID 357; cp6723>:

51 EAAYNNAIEE GKPGILVFFS ERPTPEFADL TNGSFSLSLTP IAKGFNVVVL
 101 CGPLISPLDF FHKMDPVILY MGSFLEMFPE VEAVGPRLC YILIDEQGGA
 151 QCQAVLPLET KN*

The cp7268 nucleotide sequence <SEQ ID 352> is:

5 1 ATGATGCACC GTTATTTAT TCCTTTATTA GCACCTCTCA TTTCTCTCC
 51 TTCTTAGTC AGGGCAGAGC TACAACCAAG TGAAACAGA AAAGGGGGGT
 101 GGCCTACACA ACTTCCCTGT GCAGAAGGTT CGCAACTCTT CTGAAATTC
 151 GAAGCTGCCT ATAATAATGC AATTGAGGAA GGAAACCTG GGATTTAGT
 201 CTTTTCTCT GAGCGACCCA CACAGAATT TGCCGACTTA ACGAATGGTT
 251 CATTCTCTCT CTCTACGCCA ATCGCCAAGG GCTTTAATGT CGTTGTGTTA
 301 TGCCCCGGGC TTATCAGTCC CTTAGACTTT TTCCACAAA TGGATCCTGT
 351 GATTCTCTAT ATGGGAAGTT TTCTAGAGAT GTTCCCTGAA GTGGAGGCAG
 401 TTAGTGGCCC TCGCTTATGT TATATCTAA TAGATGAACA GGTTGGGCT
 451 CAATGTCAGG CTGTCCTGCC TTTAGAAACA AAGAATTAG

15 The PSORT algorithm predicts inner membrane (0.1235).

The following *C.pneumoniae* protein (PID 4377375) was also expressed <SEQ ID 353; cp7375>:

1 1 MQRIIVGID TGVGKTIVSA ILARALNAEY WKPIQAGNLE NSDSNIVHEL
 51 SGAYCHPEAY RLHKPLSPHK AAQIDNVSIE ESHICAPKTT SNLIIETSGG
 101 FLSPCTSKRL QGDVFSSWSC SWIILVSQAYL GSINHTCLTV EAMRSRNLNI
 151 LGMVNVNGPE DEEHWLTOEI KLPIIGTLAK EKEITKTIIS CYAEQWKEVV
 201 TSNHQGIQGV SGTPSLNLH*

The cp7375 nucleotide sequence <SEQ ID 354> is:

1 1 ATGCAACGTA TCATCATTGT AGGAATCGAC ACTGGCGTAG GAAAAACCAT
 51 TGTCACTGCT ATCCTGCTA GAGCACTTAA CGCAGAACATA TGGAAACCTA
 101 TACAAGCAGG GAATCTAGAA ATTTCAGATA GCAATATTGT TCATGAGCTA
 151 TCGGGAGCCT ACTGTCTCC CGAACGTTAT CGATTGCATA AGCCCTTGTC
 201 TCCACACAAG GCAGCGAAA TCGATAATGT AAGTATCGAA GAGAGTCATA
 251 TTTGTGCGCC AAAAACAACT TCGAATCTGA TTATTGAGAC TTCAGGAGGA
 301 TTTTATCCC CCTGCACATC AAAAGACTT CAGGGAGATG TGTGTTCTTC
 351 TTGGTCATGT TCTTGGATT TAGTGAGCCA AGCATATCTC GGAAGTATCA
 401 ATCACACCTG TTTAACGGTA GAAGCAATGC GCTCACGAAA CCTCAATATC
 451 TTAGGTATGG TGGTAAATGG GTATCCAGAG GACGAAGAGC ACTGGCTAAC
 501 TCAAGAAATC AAGCTTCCCTA TAATCGGGAC TCTTGCCAAG GAAAAAGAAA
 551 TCACAAAGAC AATCATAAGC TGTATGCCG AACAAATGGAA GGAAGTATGG
 601 ACAAGCAATC ATCAGGGAAAT TCAGGGTGT A TCTGGCACCC CTTCACTCAA
 651 TCTGCATTAG

The PSORT algorithm predicts cytoplasm (0.0049).

The following *C.pneumoniae* protein (PID 4377388) was also expressed <SEQ ID 355; cp7388>:

1 1 MQVLLSPQLP PPPQHSVGSI SSPSKLRVLA ITFLVFGMLL LISGALFLTL
 51 GIPGLSAAIS FGLGIGLSAL GGVLMISGLL CLLVKREIPT VRPEEPIPEGV
 101 SLAPSEEPAL QAAQKTLAQL PKELDQLDTD IQEVFACLRK LKDSKYESRS
 151 FLNDAKKELR VFDFVVVEDTL SEIFELRQIV AQEGWDLNFL INGGRSLMMT
 201 AESESDDLHF VSKRIGLYLPS GDVRGEGLKK SAKEIVARLM SLHCEIHKVA
 251 VAFDRNSYAM AEKAFAKALG ALEESVYRSL TQSYRDKFLE SERAKIPWNG
 301 HITWLRDDAK SGCAEKKLRD AEERWKKFRK AVFWVEEDGG FDINNNLLGDW
 351 GTVLDPYRQE RMDEITFHEL YEKTTFLKRL HRKCALAKTT FEKKRSKKNL
 401 QAVEEANARR LKYVRDWYDQ EFQKAGERLE KLHALYPEVS VSIRENKIQE
 451 TRSNLEKAYE AIEENYRCCV REQEDYWKEE EKREAEFRER GNKILSPEEL
 501 ESSLEQFDHG LKNFSEKLME LEGHILKLQK EATAEVENKI LSDAESRLEI
 551 VFEDVKEMPC RIEEIEKTLR MAELPLLPTK KAFEKACSQY NSCAEMLEKV
 601 KPYCKESLAY VTSKERLVSL DEDLRRAYTE CQKRFQGDSG LESEVRACRE
 651 QLRERIQEFE TQGLDLIVEKE LLCVSSRLRN TECDCVSGVK KEAPPGKKFY
 701 AQYYDEIYRV RVQSRWMTMS ERLREGVQAC NKMLKAGLSE EDKVLKEEEY
 751 WLYREERKNK EKRLVGTIV ATQQRVAAFE SIEVPEIPEA PEEKPSLLDK
 801 ARSLFTREDH T

The cp7388 nucleotide sequence <SEQ ID 356> is:

1 1 ATGCAAGTAC TTCTATCTCC GCAGCTACCC CCCCCCCCCCC AACACTCTGT
 51 AGGGTCGATT TCTTCTCCAT CTAAACTTCG CGTTTTAGCG ATTACTTTTT

Example 181 ,
Example 182 ,
Example 183 ,
Example 184 and
5 Example 185

The following *C.pneumoniae* protein (PID 4376301) was expressed <SEQ ID 361; cp6301>:

```

1  LNQDLQNVYQ ECQKATGLES EVSAYRDHLR EQITEFETQG LDVIKEELLF
51  VSSTLKSMLS YDPLIADIPC MKFYEEYYDG IDKARVQSRW LEKSERYRKA
10  KKGFOEMLKE GLFKEDQALK KAEYRLLREK RMNKEKLIC NKIEAAQQRV
151 QEFGPSDS*

```

The cp6301 nucleotide sequence <SEQ ID 362> is:

```

1  TTGAATCAGG ATTTACAAA TGTATACCAA GAGTGCCAGA AGGCTACAGG
51  TTTAGAACG GAAGTGAGTG CATATAGAGA TCATCTTAGA GAGCAGATCA
10  CAGAGTTGA AACTCAAGGG CTGGACGTGA TAAAAGAAGA ACTTCTTTT
15  151 GTGAGTAGTA CTCTAAAG TAAATTGAGC TATGATCCAT TAATAGCAGA
201 CATTCCCTGT ATGAAGTTT ATGAGGAGTA TTATGATGGC ATTGATAAAG
251 CGAGAGTTCA ATCCCAGTGG CTGGAGAAGT CTGAGAGGTA TAGAAAGGCG
301 AAGAAGGGAT TCCAAGAGAT GCTGAAGGAA GGCCTATTCA AAGAAGATCA
351 GGCTTGAAA AAAGCAGAGT ATAGATTACT TCGAGAGAAG AGAATGAATA
401 AGGAGAAGCT TTTGATTGCA AATAAGATAG AAGCAGCTCA GCAGCGAGTC
451 CAAGAATTG GACCCTCGGA TTCATAA

```

The PSORT algorithm predicts cytoplasm (0.4621).

The following *C.pneumoniae* protein (PID 4376558) was also expressed <SEQ ID 363; cp6558>:

```

1  MNIPAPQVPV IDEPVVNNTS SYGLSLKSSL RPITYLILAI LAIATLMSVL
51  YFCGIISVGT FVLGMLIPLS VCSVLCVAYL FYQQSSIEKT KVFSITSPSV
101 FFSDEDLNLL LGREEDSVSA IDELLKNFPA DDFRRPKMLP YSNFLDEQGR
151 PNESREEDSH TSKIL*

```

The cp6558 nucleotide sequence <SEQ ID 364> is:

```

1  ATGAACATAC CCGCTCCCCA AGTACCAAGTC ATAGATGAGC CTGTAGTGAA
51  CAACACAAGT AGCTATGGTC TTTCATTGAA AAGTAGTTA AGACCGATTA
101 CTTATTGAT TTTAGCTATC TTAGCTATAG CCACACTGAT GTCTGTTCTC
151 TACTTTGTG GCATCATTAG TGTGGGACG TTGTTTTGG GCATGCTGAT
201 CCCCTCTATCG GTCTGCTCTG TTCTTGCCT TGCCATTAA TTCTATCAGC
251 AATCTTCTAT AGAAAAGACT AAGGTCTTT CTATAACCAAG TCCTTCAGTA
301 TTTTCTCTG ATGAGGATCT TAATTACTC TTAGGTGAG AAGAAGATTC
351 AGTGTCTGCA ATTGATGAAC TTCTTAAGAA CTTTCCAGCT GATGATTTC
401 GTAGGCCGAA GATGCTCCT TATTCAAATT TTCTAGATGA GCAGGGAAAGG
451 CCTAATGAGA GTAGGGAAAGA AGACTCTCAT ACTTCCAAGA TCTTATAA

```

The PSORT algorithm predicts inner membrane (0.4630).

40 The following *C.pneumoniae* protein (PID 4376630) was also expressed <SEQ ID 365; cp6630>:

```

1  MSMTIVPHAL FKNHCECHST FPLSSRTIVR IAIASLFCIG ALAALGCLAP
51  PVSYIVGSVL AFIAFVILSL VILALIFGEK KLPPTPRIIP DRFTHVIDEA
101 YGLSISAFVR EQQVTLAEFR QFSTALLCNI SPEEKIKQLP SELRSKVESF
151 GISRLAGDLE KNWNWPIFEDL LSQTCPLYWL QKFISAGDPQ VRDLGVPRE
201 CYGYYWLGPL GYSTAKATIF CKETHHILQQ LTKEDVLLK NKALQEKWDT
251 DEVKAIVERI YTTYTARGTL KTEAGGLTKE TISKELLLS LHGYSFDQLO
301 LITQLPRDAW DWLCFVDNST AYNLQLCALV GALSSQNLLD ESSIDFDVNL
351 GLYVIQDLKE AVQAFSASDE PKKELGKFLL RHLSSVSKRL ESVLRQGLHR
401 IAlehgnara RVYDVNFVTG ARIHRKTSIF FKD*

```

50 The cp6630 nucleotide sequence <SEQ ID 366> is:

```

1  ATGAGCATGA CGATCGTTCC ACATGCTTTA TTTAAAAATC ATTGCGAGTG
51  TCATTCTACC TTTCTTTGA GTTCAAGGAC TATTGTAAGA ATAGCCATTG
101 CCAGCCTCTT TTGTATAGGT GCATTAGCAG CTTTAGGCTG TTTGGCTCCT
151 CCCGTTCTTT ATATMGTGG GAGTGTGTTTA GCTTTTATTG CCTTTGTCAT
201 TCTTCTTTA GTAATTAG CTTGATTT TGGAGAGAAG AAGCTTCCAC

```

-181-

5 1 MATSVAPSPV PESSPLSHAT EVLNLPNAYI TQPHPIPAAP WETFRSKLST
 51 KHTLCFALTLL LLTLGGTISA GYAGYTGNWI ICGIGLGIIV LTLILALLA
 101 IPLKNKQTGT KLIDEISQDI SSIKGCFVQR YGLMFSTIKS VHLPELTQN
 151 QEKTRILNEI EAKKESIQNL ELKITECQNK LAQKQPKRKS SQKSFMRNSIK
 201 HLSKNPVLIF DC*

The cp6723 nucleotide sequence <SEQ ID 358> is:

10 1 ATGGCAACTT CCGTAGCCCC ATCACCAAGTC CCCGAGAGCA GCCCTCTCTC
 51 TCATGCTACA GAAGTTCTCA ATCTTCCTAA TGCTTATATT ACGCAGCCTC
 101 ATCCGATTCC AGCGGCTCT TGGGAGACCT TTGCTCCAA ACTTTCCACA
 151 AAGCATAACGC TCTGTTTGC CTAAACACTA CTGTTAACCT TAGGGGGAAC
 201 GATCTCAGCA GGTTACGCAG GATATACTGG AAACCTGGATC ATCTGTGGCA
 251 TCGGCTTGGG AATTATCGTA CTCACACTGA TTCTTGCTCT TCTTCAGCA
 301 ATCCCCTTTA AAAATAAGCA GACAGGAACA AAACCTGATTC ATGAGATATC
 351 TCAAGACATT TCCTCTATAG GATCAGGATT TGTTCAAGAGA TACGGGTTGA
 401 TGTTCTCTAC AATTAAAAGC GTGCATCTTC CAGAGCTGAC AACACAAAAT
 451 CAAGAAAAAA CAAGAATTAA AAATGAAATT GAAGCGAAAA AGGAATCGAT
 501 CCAAAATCTT GAGCTTTAAA TTACTGAGTG CCAAAACAAAG TTAGCACAGA
 551 AACAGCCGAA ACGGAAATCA TCTCAGAAAT CATTATGCG TAGTATTAAG
 601 CACCTCTCCA AGAACCCCTGT AATTGCTTC GATTGCTGA

20 The PSORT algorithm predicts inner membrane (0.6095).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 179A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 179B) and for FACS analysis.

25 These experiments show that cp6723 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 180

The following *C.pneumoniae* protein (PID 4376749) was expressed <SEQ ID 359; cp6749>:

30 1 MSYYFSLWYL KVQQHFQAAF DFTRSLCSRI SNFALGVIAL LPIIGQLYVG
 51 LDWLLSRIKK PEFPSDVQI VRVEHVVGHM HRSRVEDILK RQRLSLEPRD
 101 EGKVHGDLPS APFF*

The cp6749 nucleotide sequence <SEQ ID 360> is:

35 1 ATGAGTTATT ACTTTCTCT TTGGTATCTG AAGGTGCAAC AGCACTTCA
 51 AGCAGCATTG GATTTACTC GCTCCCTGTG TTCACGAATT TCTAATTTTG
 101 CTTTGGGAGT GATTGCATTG CTTCCCTATTA TTGGGCAGTT GTATGTAGGG
 151 CTGGACTGGC CCCTCTCTAG GATAAAAAAG CCAGAATTTC CTTCCGATGT
 201 GGATCAGATC GTGCGAGTAG AACACGTCGT GGGTCACGAC CATAGAAGTC
 251 GAGTTGAAGA TATTCTAAAG AGACAAAGGC TCTCATTAGA GCCTAGAGAC
 301 GAGGGGAAGG TTCACGGAGA TCTGCCTTC GCTCCTTTT TTTGA

The PSORT algorithm predicts inner membrane (0.2996).

40 The protein was expressed in *E.coli* and purified as a his-tag product (Figure 180A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 180B) and for FACS analysis.

These experiments show that cp6749 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

5

```

51 CACTAAAAAA AAATCCTGCA GCAACTTTGA TAAGATTCAG TCTCGAATTG
101 TATTGATTAC TGCAATCTT GCTGTCCTTAG TTACTATAGG GACCCTACTT
151 ATTTGGTTTGCA TTTTAATAT TCCGTGTTATC TATTCCTCTA CAGGAATTTC
201 ATTTTATTGCT GTTGTCTTCA GCAACTTTAT CCTTTATAAA CGAGCAACCA
251 CCCTCTTAAA ACCGCGTGCT TGTGGCAAAC ACAAAAGAAAT AAAACCAAAA
301 AGGGTCTCCA CCAACCTACA GTATTCTTCT ATCTCTATCG CAATCAATCG
351 TTCTAAAGAA AACTGGGAAC ACCAACCCAA GGACCTACAG AATCTCCCCG
401 CACCCCTCTGC ATTACTCACA GATAACCCCT ACAGAGATATG GAAAGCTAAA
451 CATTCACTGT TTTCCCTAGT ATCCCTCTA CGGGGAGGCA ATCCAGAAC
501 TCTCTTAAATT TCAGCTTCCG AAAATTAGG AAAGACTCTG TTAATTGAAG
551 AAACCTCGCA AAATGCCCT ATATCCTCCT ACCTAGATAC CACTCCCTCC
601 CCAAAATCCT TGCTCAATGA GGCAATTCAAG GAAACCCAGGG TAGAAATAAA
651 TACAGAACTC CCTGCGGGAG ATTCAAGGAGA ACGTTTATAC TGGCAACCCG
701 ATTCCCGAGG CGCGCTCTTC CTCCCACAAA TACCAACAAAC TCCTGAAGCC
751 ATCTACCAAT ACTACTATGC ACTCTATGTC ACTTATATAC AGACTGCGAT
801 CAATACGAAC ACCCAAATTA TCCAAATCCC TTATACAGC TTGAGGGAGC
851 ATCTCTATTG TAGAGAATTG CCCCCGCAAT CAAGAATGCA ACAATCTTG
901 GCTATGATTAA CAGCAGTAAA ATACATGGCC GAGCTGCACC CAGAATATCC
951 GCTAACTATTG GCTTGTTG AAAGATCCTT AGCCCAACTA CCTCAAGAAA
1001 GTATTGAGGA TCTCTCTTAG

```

The PSORT algorithm predicts inner membrane (0.5288).

The proteins were expressed in *E.coli* and purified as GST-fusion products. The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 181-185) and for FACS analysis.

25 These experiments show that cp6301, cp6558, cp6630, cp6633 and cp6642 are surface-exposed and immunoaccessible proteins, and that they are useful immunogens. These properties are not evident from their sequences alone.

Example 186

The following *C.pneumoniae* protein (PID 4376389) was expressed <SEQ ID 371; cp6389>:

30

```

1 MSEVKPLFLK NDSFDLATORF QNLINMLQE QAEIYNEYEE KNARVQNEIK
51 EQKDFVKRCL EDFEARGLGV LKEELASLTR DFHDKAKAET SMLIECPCIG
101 FYYSIHQEQQ RQRQERLQKM AERYRDCKQV LEAVQVEQKD MISSRVVVDD
151 SYFEEEKEEQ KVDRNRKKEQD *

```

The cp6389 nucleotide sequence <SEQ ID 372> is:

35

```

1 ATGTCAGAAG TGAAGCCTTT GTTTTAAAG AATGACTCTT TTGATTTGGC
51 AACTCAGAGA TTCCAGAACAT TAATTAACAT GCTACAAGAG CAAGCCGAGA
101 TATATAACGA GTATGAAGAA AAGAATGCTA GGGTTTCAGAA TGAGATTAAG
151 GAGCAAAAGG ACTTTGTGAA AAGATGCATA GAGGACTTTG AAGCCAGAGG
201 ACTGGGGGTG CTAAAAGAAG AGCTTGCATC TTTGACGCGT GATTTCATG
251 ATAAAGCAAA AGCAGAGACT TCTATGCTCA TTGAATGTCC TTGTATTGGT
301 TTTTATTATA GTATTCACTCA GGAGGAACAA AGGCAAAGGC AAGAAAGGCT
351 TCAAAAGATG GCTGAGCGCT ATAGGGACTG TAAACAAAGTC TTGGAGGCTG
401 TCCAGGTGGA GCAGAAAGAT ATGATATCTT CTAGAGTCGT TGTGATGAC
451 AGCTACTTTG AAGAAGAAAA AGAAGAACAA AAGGTGGATA ACAGAAAGAA
501 AGAACACAGGAC TAG

```

The PSORT algorithm predicts cytoplasm (0.3193).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 186A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 186B) and for FACS analysis.

251 CAACACCAAG AATCATCCT GATAGATTAA CTCACGTGAT AGATGAAGCT
 301 TATGGCCTTT CAATCTCTGC ATTTGTAAGA GAACAGCAGG TAACATTAGC
 351 CGAGTTTAGA CAATTTCTCA CTGCCCTGTT GTGTAACATA TCTCCTGAAG
 401 AGAAAATCAA ACAATTGCCT TCTGAATTGC GAAGTAAAGT AGAGAGTTTT
 451 GGTATTAGCA GGCTCGCAGG TGAGTTAGAA AAGAATAATT GGCCAATATT
 501 TGAAGATCTT TTAAGCCAAA CCTGCCGTT ATATTGGCTT CAGAAATTAA
 551 TATCAGCAGG AGATCCACAA GTTGTAGAG ACCTAGGTGT CCCTAGAGAA
 601 TGTATGGGT ACTATTGGCT AGGGCCTTG GGATACAGTA CAGCTAAGGC
 651 TACAATTTT TGTAAAGAGA CGCATCATAT TCTTCAACAA TAAACGAAAG
 701 AGGACGTTCT TTTATTAAGA AACAAGGCTC TTCAAGAGAA ATGGGATACT
 751 GATGAAGTCA AAGCAATTGT AGAGCGTATC TACACTACCT ATACGGCACG
 801 AGGAACCTCA AAGACCGAAG CAGGGGGACT TACAAAAGAG ACAATCAGTA
 851 AGGAATTGCT ATTGTTGAGC TTGCATGGCT ATTCTTTGA TCAGCTACAG
 901 CTGATCACTC AACTTCCCTAG AGATGCTTGG GATTGGCTGT GTTTTGTAGA
 951 TAACAGTACC GCATACAAAC TTCAGCTTG TGCTCTTGTG GGAGCTTTGT
 1001 CATCCCCAAA TCTTCTGAC GAATCTTCTA TCAGATTTGA TGTAACACTA
 1051 GGCCCTGTATG TGATTCAGGA TCTAAAGAA GCTGTTCAAG CATTTCCTGC
 1101 TTCTGATGAG CCAAAGAAAG AACTAGGTAA ATTCTTGTGTA AGGCATTTGA
 1151 GTTCAGTTTC TAAGCGATTA GAGAGTGTAT TAAGACAGGG TCTTCACAGA
 1201 ATAGCTCTAG AGCATGGAAA TGCCAGAGCT AGGGTTTATG ACGTCAATT
 1251 TGTAACAGGA GCTAGAACATTG ATAGGAAGAC GAGTATCTTC TTTAAAGACT
 1301 AA

The PSORT algorithm predicts inner membrane (0.7092).

The following *C.pneumoniae* protein (PID 4376633) was also expressed <SEQ ID 367; cp6633>:

25 1 MVNIQPVYRN TQVNYSQATQ FSVCQPALSL IIVSVVAAL AIVALVCSQS
 51 LLSIELGTAL VLVSLILFAS AMFMITYKMRQ EPKEELLIPKK IMELIQEHYP
 101 SIVVDFIRDQ EVSIYEIHHL ISILNKTNVF DKAPVYIQLQEK LLQFGIEKFK
 151 DVHPSKLPNF EEILLQHCPL HWLGLRVYPM VSDVTPGTYG YYWCGPLGLY
 201 ENAPSLFERR SLLLKKISF GEFALEEDGL KKNTWSSSEL VQIRQNLFTR
 251 YYADKEEVDE AELNADYEQF DSSLHLIFSH KLS*

The cp6633 nucleotide sequence <SEQ ID 368> is:

35 1 ATGGTTAATA TACAGCCTGT GTATAGGAAT ACCCAAGTCA ACTATAGTCA
 51 GGCTACCAA TTTTCGGTGT GCCAGCCAGC GCTTAGCCTG ATTATCGTTT
 101 CTGTTGTTGC TGCTGTACTC GCTATTGTAG CTTTGGTAG CAGTCATCT
 151 CTTTTATCCA TAGAGTTAGG AACTGCTCTT GTTCTAGTTT CTCTTATTCT
 201 TTTGCTTCT GCTATGTTA TGATTTATAA GATGAGACAA GAACCTAAGG
 251 AGTTGCTGAT CCCTAAGAAA ATCATGGAAC TCATCCAAGA ACATTATCCA
 301 AGTATTGTTG TTGATTTTAT TAGAGATCAG GAGGTTTCCA TTTATGAGAT
 351 ACATCACTTC ATCTCTATTG TTAATAAGAC GAATGTTTC GACAAAGCAC
 401 CAGTATATTG ACAAGAAAAA CTCTTACAGT TTGGCATTTGA GAAGTTCAAA
 451 GATGTACATC CAAGTAAGCT CCCTAATTTC GAAGAAATTG TTCTACAGCA
 501 TTGCCCATTG CATTGGTTGG GACGTCTGGT ATATCCCCTG GTATCGGATG
 551 TCACTCCAGG AACCTATGGA TACTATTGGT GTGGTCCTT AGGACTGTAC
 601 GAGAACGCTC CCTCTCTTT TGAAACGTGCA TCTCTTCTAT TGTTAAAGAA
 651 AATTAGCTTT GGAGAGTTG CTCTTTTAGA AGATGGTCTC AAGAAAAACA
 701 CGTGGAGTTG TTCGGAACTC GTTCAAATCA GACAAACCT TTTTACAAGA
 751 TATTATGCTG ATAAAGAAGA GGTAGATGAA GCAGAGTTAA ACGCTGATTA
 801 CGAACAGTTT GATTCCCTCC TTCACCTTAT TTTTCTCAC AAGCTCTCTT
 851 GA

50 The PSORT algorithm predicts inner membrane (0.7283).

The following *C.pneumoniae* protein (PID 4376642) was also expressed <SEQ ID 369; cp6642>:

55 1 MATISPISLT VDHPLVDTKK KSCSNFDKIQ SRILLITAIF AVLVTIGTLL
 51 IGLLLNIPVI YFLTGISFIA VVLSNFIYK RATLLKPRCA CGKHKEIKPK
 101 RVSTNLQYSS ISIAINSKE NWEHQPKDLQ NLPPAPSALLT DNPyEIWKAK
 151 HSLFSLVSLL PGGNPEHLLI SASENLGKTL LIEETSQNAP ISSYVDTPPS
 201 PKSLLNEAIQ ETRVEINTEL PAGDSGERLY WQPDFRGRVF LPQIPTTPEA
 251 IYQQYYALVV TYIQTAINN TQIWIPLYS LREHLYSREL PPQSRMQQSL
 301 AMITAVKYM ELHPEYPLTI ACVERSLAQL PQESIEDLS*

The cp6642 nucleotide sequence <SEQ ID 370> is:

60 1 ATGGCTACAA TCTCACCCAT ATCTTTAACT GTAGATCATC CCCTAGTAGA

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 188A; lanes 2-3). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 188B) and for FACS analysis.

These experiments show that cp6868 is a surface-exposed and immunoaccessible protein, and that it
5 is a useful immunogen. These properties are not evident from the sequence alone.

Example 189

The following *C.pneumoniae* protein (PID 4376894) was expressed <SEQ ID 377; cp6894>:

10	1	MYKRCVLDKI	LKGIVAGSLI	LLYWSSDLLE	RDIKSIGKGV	RDIQEDIREI
	51	SRVVVKQQTS	QAIPAAPGVM	LAPKLRDEA	FALLFGDPSY	PNLSSLDPYK
	101	QQTLPPELLGT	NFPHGILRT	AHVGPENLS	PFNGFDYVVG	FYDLCIPS LA
	151	SPHVGKYEEF	SPDLAVKIEE	HLVEDGSGDK	EFHIYLPRNV	FWRPIPDPKAL
	201	PKHVQLDEVF	QRPHPVTAHD	IKFFYDAVMN	PYVATMRAVA	LRSCYEDVVS
	251	VSVENDLKL V	VRWKAHTVIN	EEGKEERKV L	YSAFSNTLSL	QPLPRFVYQY
	301	FANGEKIIIED	ENIDTYRTNS	IWAQNFTMHW	ANNYIVSCGA	YYFAGMDDEK
15	351	IVFSRNPDY	DPLAALIDKR	FVFYKESTDS	LFQDFKTGKI	DISYLPNQR
	401	DNFYSFMKSS	AYNKQVAKGG	AVRETVSADR	AYTYIGWNCF	SLFFQSRQVR
	451	CAMNMAIDRE	RIIEQCLLDQ	GYTISGPFA S	SSPSYNNQIE	GWHYSPEEAA
	501	RLLEEEGWID	TDGDGIREKV	IDGVIVPFR	RLCYYVKS VT	AHTIADYVAT
20	551	ACKEIGIECS	LLGLDMA DLS	OAFDEKNFDA	LLMGWCLGIP	PEDPRALWHS
	601	EGAMEKGSAN	VVGPHNEEAD	KIIDRLSYEY	DLKERNRLYH	RFHEIIHEEA
	651	PY AFLFSRHC	SLLYKD YVKN	IFVPTHRTDL	IPEAQDET VN	VTMVWLEKKE
	701	DPCLSTS*				

The cp6894 nucleotide sequence <SEQ ID 378> is:

25	1	ATGTATAAAA	GATGTGTGCT	AGATAAAATT	TTAAAGGGGA	TTGTCGCCGG
	51	TTCTTTAATT	TTGTTATACT	GGTCCTCAGA	CCTACTTGAA	AGAGACATTA
	101	AGTCGATAAA	AGGTAACGTA	AGAGATATT C	AAGAACAT	TCTGAAATC
	151	TCA CGCGTAG	TGAAACAACA	GCAGACATCA	CAAGCTATCC	CTGCGGCACC
	201	TGGGGTGATG	CTCGCTCCTA	AGCTCGTCAG	AGACGAAGCT	TTTGCTCTAC
	251	TCTTTGGAGA	TCCTAGTTAT	CCTAATTTAC	TTTCCCTAGA	CCCCTATAAA
30	301	CAGCAGACTC	TTCCTGA ACT	TCTAGGAACA	AATTTCACC	CTCATGGTAT
	351	CCTACGCACT	GCCCCATGTCG	AAAAACCCGA	AAATCTGAGC	CCTTTTAATG
	401	GCTTGATT A	TGTCGTGGC	TTTACGATC	TCTGTATTCC	TAGTTTAGCT
	451	TCTCCCCACG	TAGGGAAATA	CGAAGAATT T	TCTCCAGATC	TCGCTGTGAA
35	501	AATAGAAGAA	CATCTGTTG	AAGATGGTC	TGGGGATAAA	GAGTTTCACA
	551	TCTATCTGAG	GCCGAATGTT	TTTGGCGTC	CTATAGATCC	TAAGGCCCTT
	601	CCAAAACACG	TTCAGTTAGA	CGAAGTATT T	CAACGTCCTC	ATCCTGTGAC
	651	AGCTCATGAT	ATTAAGTTT	TCTACGACGC	TGTTATGAAC	CCTTATGTAG
	701	CAACCATGCG	AGCAGTGGCT	CTGCGCTCTT	GTTATGAAGA	TGTGGTTCT
40	751	GTC TCA GTAG	AAAACGATT T	AAAATTAGTA	GTCAGATGGA	AAGCACACAC
	801	GGTAATCAAT	GAAGAAGGAA	AGGAAGAGCG	CAAAGTGC TC	TACTCTGCAT
	851	TTTCTAATAC	CTTAAGCTTG	CAGCCCCCTCC	CTAGATTGT	ATATCAGTAT
	901	TTTGCTAACG	GGGAAAAAAT	CATTGAAGAT	GAGAATATCG	ATACCTACCG
45	951	AACCAATTCC	ATTTGGCGC	AAAACCTCAC	TATGCATTGG	GCAAACAAC T
	1001	ATATTGTAAG	TTGTGGAGC	TACTACTTG	CAGGGATGGA	TGATGAGAAA
	1051	ATCGTGT TTT	CTAGAAATCC	TGACTTCTAT	GATCCCTCTG	CGGCTCTTAT
	1101	TGACAAGCGT	TTCGTCTATT	TTAAGGAAAG	CACAGACTCC	CTATTCCAAG
	1151	ATTTTAAGAC	AGGGAAAATA	GACATCTCTT	ACCTTCACC	CAACCAAAGA
50	1201	GATAATTCT	ATAGTTTAT	AAAAGCTCC	GCTTATAACA	AACAGGTAGC
	1251	TAAGGGAGGA	GCCGTCCGTG	AAACAGTCTC	AGCAGATCGA	GCATATA CGT
	1301	ACATAGGATG	GAATTGCTTT	TCATTATTT	TCCAAAGCCG	ACAGGTGC GC
	1351	TGTGCTATGA	ACATGGCAAT	CGATAGAGAG	AGGATTATCG	AACAGTGCTT
	1401	GGATGGCCAA	GGCTATACGA	TTAGTGGGCC	TTTGCTCTG	AGTTCTCCTT
	1451	CTATAATAA	ACAGATCGA	GGGTGGCATT	ATTCTCCAGA	AGAAGCAGCT
55	1501	CGTCTCCTGG	AAGAAGAGGG	ATGGATAGAT	ACCGATGGCG	ATGGAATCCG
	1551	AGAAAAAAGT	ATCGATGGTG	TGATTGTCCC	GTTCCGTTTC	CGTTTATGCT
	1601	ATTATGTA A	GAGTGTCA CC	GCTCATACCA	TTGCAGATTA	CGTAGCTACT
	1651	GCTTGTAAAG	AAATCGGAAT	CGAGTGTAGC	CTTCTAGGAC	TAGATATGGC
	1701	CGATCTTCC	CAAGCTTTG	ATGAAAAGAA	TTTCGATGCT	CTTTTTATGG
	1751	GATGGTGT T	AGGAATTCC	CCTGAGGATC	CTAGGGCTTT	ATGGCATTCT

These experiments show that cp6389 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 187

The following *C.pneumoniae* protein (PID 4376792) was expressed <SEQ ID 373; cp6792>:

```

5      1  VLQEHFFLSE DVITLAQQLL GHKLITTHEG LITSGYIVET EAYRGPDCKA
      51 CHAYNYRKTO RNRAMYIKGG SAYLYRCYGM HHLLNVVTGP EDIPHAVLIR
     101 AILPDQGKEL MIQRQRWRDK PPHLLTNGPG KVCQALGISL ENNRQLNTP
     151 ALYISKEKIS GTLTATARIG IDYAQEYRDV PWRFLLSPED SGKVL*
```

The cp6792 nucleotide sequence <SEQ ID 374> is:

```

10     1  GTGCTACAAG AACATTTTT TCTATCGGAA GATGTAATT CACTAGCGCA
      51 ACAGCTTTA GGACATAAAC TCATCACAAC ACATGAGGGT CTGATAACTT
     101 CAGGTTACAT TGTAGAAACC GAAGCGTATC GTGGCCCTGA TGACAAAGCA
     151 TGCCACGCCT ACAACTACAG AAAAACTCTAG AGGAACAGAG CGATGTACCT
     201 GAAAGGAGGC TCTGCTTACC TCTACCGTTG CTATGGCATG CATCACCTAT
     251 TGAATGTGTC CACTGGACCT GAGGACATT CCCATGCCGT CCTGATCCGG
     301 GCCATCCTTC CTGATCAAGG CAAAGAACTT ATGATCCAAC GCGGCCAATG
     351 GAGAGATAAA CCCCCACACC TTCTCACCAA TGGACCCGGAA AAAGTGTGCCC
     401 AAGCTCTAGG AATCTCTTG GAAAACAATA GGCAACGCCT AAATACCCCA
     451 GCTCTCTATA TCAGCAAAGA AAAAATCTCT GGGACTCTAA CAGCAACTGC
     501 CCGGATCGGC ATCGATTATG CTCAAGAGTA TCGTGATGTC CCATGGAGAT
     551 TTCTCCTATC CCCAGAAAGAT TCGGGAAAAG TTTTATCTTA A
```

The PSORT algorithm predicts cytoplasm (0.180).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 187A; lanes 2-4). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 187B) and for FACS analysis.

These experiments show that cp6792 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 188

The following *C.pneumoniae* protein (PID 4376868) was expressed <SEQ ID 375; cp6868>:

```

30     1  MVETVLHNFQ RYLSKYLYRV FRFPCRKKTF LSSHRVLARP SFPVDYCPGK
      51 IYDLQEYEE LNAQLFQGAL RLQIGWFGRK ATRKGKSVVL GLFHENEQLI
     101 RIHRSLDRQE IPRFFMEYLV YHEMVHSVP REYSLSGRSI FHGKKFKEYE
     151 QRFPLYDRAV AWEKANAYLL RGYKKRVGCGG YGRA*
```

The cp6868 nucleotide sequence <SEQ ID 376> is:

```

35     1  ATGGTTGAAA CAGTACTTCA TAATTTCCAA CGTTATCTGA GCAAGTATCT
      51 CTATAGGGTA TTTCGCTTCC CATGTCGTA AAAGACGTTT CTATCTTCGC
     101 ACAGGGTTCT TGCTCGTCCT TCATTCCCAG TAGACTACTG TCCGGGAAAG
     151 ATCTATGATT TGCAGGAGAT CTATGAGGAA TTGAATGCGC AGTTATTTC
     201 AGGTGCACTG CGTTTACAGA TTGGTTGGTT CGGAAGGAAA GCTACCAGAA
     251 AAGGCAAGAG TGTTGTCTTG GGATTGTTTC ATGAAAATGA ACAGTTAATT
     301 CGAATTCACTC GTTCTTTAGA TCGGCAGGAA ATCCCAAGAT TTTTTATGGA
     351 ATATCTTGTG TATCATGAAA TGGTTCATAG TGTAGTCCT AGAGAGTATT
     401 CTCATCGGG GCGTTCGATT TTTCATGGTA AAAAGTTAA AGAATACGAA
     451 CAACGTTTCC CCTTGTATGA TCGTGCTGTT GCTTGGGAAA AGGCAAACGC
     501 TTATTTATTG CGAGGGTATA AAAAAAGAGT AGGTGGAGGA TATGGCAGGG
     551 CATAG
```

The PSORT algorithm predicts bacterial cytoplasm (0.325).

TABLE II – sequences of the primers used to amplify Cpn genes.

Orf ID	N-terminus final primer	C-terminus final primer
CP0014P	GCGTC CCG GGT CATATG AAGTCTTCTTCCCCA	GCGT CTC GAG ATGAAAGAGTTTGCG
CP0015P	GCGTCCGGGTCAATATG TCAGCTCTGTTTCTGA	GCGT CTC GAG GAATTGGTATTTGCTC
CP0016P	GCGTCCGGGTCAATATG GCGATCTCACATTAG	GCGT CTC GAG GTCCAAGTAAAGTAGCA
CP0017P	GCGT CCG GGT CATATG GGTATCAAGGGAACTG	GCGT CTC GAG AAATCCGAATCTTCC
CP0019P	GCGTCCGGGTCAATATG CAAGACTCTCAAGACTATAG	GCGT CTC GAG AAATGGTATTTACCC
CP6260P	GCGTC CCG GGT GCTAGCACTACGATTTCTTAAACC	GCGT CTC GAG AAAACGAAATTGCTTC
CP6397P	GCGTC CCG GGT CATATG GTTAAACTGCTAAAAATCTATT	GCGT CTC GAG ATGAAAGAAGAGTCCCTCG
CP6456P	GCGTC CCG GGT CATATG TCATCTCTGTAAATAACA	GCGT CTC GAG CTGACCATCTCTGTT
CP6466P	GCGTC CCG GGT CAT ATG TGCAAGGAGTCCAGT	GCGT CTC GAG ATTTTCCTTACGATAACG
CP6467P	GCGTC CCG GGT CAT ATG TGTCCCCATCCCAA	GCGT CTC GAG TAGTTTTCTATAAACGAAAGTCT
CP6468P	GCGTC CCG GGT CAT ATG TGCTCTCTACTCTTC	GCGT CTC GAG GGGGAAATAGGTATATTGAA
CP6469P	GCGTC CCG GGT CAT ATG AGCTGCTAAAGCAA	GCGT CTC GAG ACTTAAGATATCGATATTGAA
CP6552P	GCGTC CCG GGT CAT ATG TGCCATAAGGAAGATG	GCGT CTC GAG ACCATTGTCCTGAGTCAT
CP6567P	GCGTC CCG GGT CAT ATG ACCTCACCGATCCCC	GCGT CTC GAG AGAACCCGTAGAGGC
CP6576P	GCGTC CCG GGT CAT ATG ACTGAAAAGTTAAAGAAGG	GCGT CTC GAG GAA CATGCCCTAA
CP6727P	GCGTC CCG GGT CATATGCTACATCCACTAATGGC	GCGT CTC GAG GAAAGAATAACGAGTTC
CP6729P	GCGTC CCG GGT CAT ATGGCAGATGCTTCTTATC	GCGT CTC GAG GAATGAGTATCTTAGCC
CP6731P	GCGTC CCG GGT CATATGCCGTGTTGAATCAAT	GCGTC CAT GGC GGC CGC GAACCTGAACTTACCTCC
CP6736P	GCGTC CCG GGT GCT AGCGTAGAAGTTATCATCCCT	GCGTC CAT GGC GGC CGC AAATCGTAATTGCTTC
CP6737P	GCGT GGA TCC CAT ATG GAGACTAGACTCGGAGG	GCGT CTC GAG AAATGTGGATTAGTCC
CP6751P	GCGTC CCG GGT GCT AGC AATGAAGGCTCCAACT	GCGT CTC GAG AAATCTCATTCTACTCGC
CP6752P	GCGTGA ATT CAT ATGTTGGGATGACTCCT	GCGT CTC GAG GAATTAAAGGTACTCCTG
CP6753P	GCGTC CCG GGT GCT AGCACTCCTACTCTCATAGAG	GCGT CTC GAG AAACCTAAAGGTGTT
CP6787P	GCGTC CCG GGT CAT ATG ATAAAACAAATAGGCCGT	GCGT CTC GAG TTCGTAAGCAACTTCAGA
CP6829P	GCGTC CCG GGT CAT ATG AAGCAGATGCGTCTT	GCGTC CAT GGC GGC CGC GAAACTAAGGGAGAGGC
CP6830P	GCGTC CCG GGT CAT ATG GATCCCCTGCTGTT	GCGTC CAT GGC GGC CGC GAATACAACCGGATCC
CP6832P	GCGTC CCG GGT CAT ATG CATAAAGTAATAGTTTCATT	GCGT CTC GAG TAAACTAGAAAAAGTCGTC
CP6848P	GCGTC CCG GGT CAT ATG TCATCAAATCTACATCCC	GCGT CTC GAG AACCGGAGCTTTTAC
CP6849P	GCGTC CCG GGT GCT AGC AGCGGGGTATAGAG	GCGT CTC GAG ATACACGTGGGATTTTC
CP6850P	GCGTC CCG GGT CAT ATG TGCCGCATTGAGAT	GCGT CTC GAG CTGTTGCATCTGCC
CP6854P	GCGTC CCG GGT GCT AGC TCAATAGCTATTGCAAG	GCGT CTC GAG TTATCGAAATGCTTTG
CP6879P	GCGTC CCG GGT CAT ATG GCAACACCCGCTCAA	GCGTC CAT GGC GGC CGC TCCTTGAATTGCTCTGC
CP6894P	GCGTC CCG GGT CAT ATG TATAAAAGATGTTGCTAGA	GCGT CTC GAG GGATGTAACCTAACGACCG
CP6900P	GCGTC CCG GGT CAT ATG AAGATAAAATTTCTTGAAG	GCGT AAG CTT GGGAAACGATACCG
CP6952P	GCGTC CCG GGT CAT ATG CTCTCGGATCAATATAGG	GCGT CTC GAG TCAGATTCTTTTTAGC
CP7034P	GCGTC CCG GGT CAT ATG AAAAAACAGGTATATCAATG	GCGT AAG CTT AAACGCTGAAATTATACC
CP7090P	GCGTC CCG GGT CAT ATG TGTAGCCTTCCCT	GCGT CTC GAG GCGTGCATGAATCTTA
CP7091P	GCGTC CCG GGT CAT ATG GAAGAATTAGAAGTTGTTG	GCGT CTC GAG TAGTGTCTCTTATCGGT
CP7170P	GCGTC CCG GGT CAT ATG CTAGGGCTGGAAACC	GCGT AAG CTT AACTCGAGACCTGACG
CP7228P	GCGTC CCG GGT CAT ATG ACTCTGTTCTTATTCTACA	GCGT CTC GAG ATCTGAAAGCGGAGG
CP7249P	GCGTC CCG GGT CAT ATG ATCCCCATCCCTTAC	GCGT CTC GAG ATCAGTTGCTGAGACTT
CP7250P	GCGTC CCG GGT CAT ATG AATCTTCAACAGGTCT	GCGT CTC GAG ATTTCCTAGAGAGACTCTC
CP0018P	GTGCGT CATATG GCAACCACTCCACTAA	ACTCGCTA GCGGCCGC TAATGAGGTCCCCAG
CP6270P	GTGCGT CATATG AATTATTAGGAGCTGCT	ACTCGCTA GCGGCCGC AAATTGATTTGCTACC
CP6735P	GTGCGT CATATG GCAGCACAAAGTTATAT	ACTCGCTA GCGGCCGC TGGCGTAGAAAGTATC
CP6998P	GTGCGT CATATG TTGCCGTAGGGAAAC	ACTCGCTA GCGGCCGC GAATCTGAACTGACCA
CP7033P	GTGCGT CATATG GTTAATCTTATGGTCCA	ACTCGCTA GCGGCCGC TTGGAGATAACCGAAATATA
CP7287P	GTGCGT CATATG TTACACAGCTCAGAACTAGA	ACTCGCTA GCGGCCGC GAAAATAATACGGATACCA
CP0010P	GTGCGT CATATG GCAACTGCTAAAAATA	GCGT CTCGAG GAATTGGAACCTACCC
CP0468P	GTGCGT GCTAGC ATTTCCTATGACAAACTCTAT	GCGT CTCGAG AAATGTGCAATGACTCT
CP6272P	GTGCGT CATATG TTGACTCATCAAGAGGCT	GCGT CTCGAG GAAGGGAGTTTTAGGT
CP6273P	GTGCGT CATATG ACATATCTGGAAGCTC	ACTCGCTA GCGGCCGC CTCCACAATTCTATG
CP6362P	GTGCGT CATATG CCCCTTGATATTACTTATTAACA	GCGT CTCGAG TCGTTCCAATCCA
CP6372P	GTGCGT CATATG AAACAAACATTCCTCTAAATA	GCGT CTCGAG TTTCTTGTGGTTTCT
CP6390P	GTGCGT CATATG CGAGAGGTGCTTAAG	ACTCGCTA GCGGCCGC TCTCTTAGACAGCCCTT
CP6402P	GTGCGT CATATG AATGTTGGGATCTCTT	GCGT CTCGAG GAAGGGGTGCCCCGT
CP6446P	GTGCGT CATATG TGTAAATCAAAGCCCTCTT	GCGT CTCGAG GGGCTGAGGAGGAAC
CP6520P	GTGCGT GCTAGC AAACACTACCTATCATTTCT	GCGT CTCGAG CAGAAAGGCTTTCTT
CP6577P	GTGCGT CATATG AATTAGGCTATGTTAATTAA	GCGT CTCGAG GTTTGTGAAAGA
CP6602P	GTGCGT CATATG GCAGCATCAGGAGGCA	GCGT CTCGAG TGACCAAGGATAGGGTTAG

5
 1801 GAAGGGGCTA TGGAAAAGGG TTCAGCGAAT GTTGTAGGTT TCCATAATGA
 1851 AGAACGCTGAT AAAATCATAG ACAGACTCGAG CTACGAATAC GATCTGAAAG
 1901 AACGTAATCG CCTGTACCAAC CGTTTCCATG AAATTATTCA TGAGGAAGCT
 1951 CCTTATGCTT TCTTGTTCTC ACGACATTGT TCCTTACTTT ATAAGGATTA
 2001 TGAAAAAAAT ATTTTCGTAC CTACACATAG AACAGATTAA ATTCCCTGAAG
 2051 CTCAGGATGA GACTGTCAAC GTAATATGG TATGGCTTGA GAAGAAGGAG
 2101 GATCCGTGCT TAAGTACATC CTAA

The PSORT algorithm predicts inner membrane (0.162).

10 The protein was expressed in *E.coli* and purified as a his-tag product (Figure 189A) and also in GST/his form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 189B) and for FACS analysis.

These experiments show that cp6894 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 190

15 The following *C.pneumoniae* protein (PID 4377193) was identified in the 2D-PAGE experiment <SEQ ID 379; cp7193>:

20
 1 MKRVVIYKTIF CGLTLLTSLS SCSLDPKGYN LETKNSRDLN QESVILKENR
 51 ETPSLVVKRLS RRSRRLFARR DQTQKDTLQV QANFKTYAEK ISEQQDERDLS
 101 FVVSSAAEKS SISLALSQGE IKDALYRIRE VHPLALIEAL AENPALIEGM
 151 KKMQGRDWIWLNFLTQLSEV FSQAWSQGVI SEEDIAAFAS TLGLDSGTVA
 201 SIVQGERWPE LVDIVIT*

A predicted leader peptide is underlined.

The cp7193 nucleotide sequence <SEQ ID 380> is:

25
 1 ATGAAAAGAG TCATTTATAA AACCATATTT TGCGGGTTAA CTTTACTTAC
 51 AAGTTTGAGT AGTTGTTCCC TGGATCCTAA AGGATATAAC CTAGAGACAA
 101 AAAACTCGAG GGACTTAAAT CAAGAGTCTG TTATACTGAA GGAAAACCGT
 151 GAAACACCTT CTCTTGTAA GAGACTCTCT CGTCGTTCTC GAAGACTCTT
 201 CGCTCGACGT GATCAAACTC AGAAGGATAC GCTGCAAGTG CAAGCTAACT
 251 TTAAGACCTA CGCAGAAAAG ATTTCAGAGC AGGACGAAAG AGACCTTTCT
 30 301 TTTCGTTGTCT CGTCTGCTGC AGAAAAGTCT TCAATTTCGT TAGCTTTGTC
 351 TCAGGGTGAA ATTAAGGATG CTTTGTACCG TATCCGAGAA GTCCACCCCTC
 401 TAGCTTTAAT AGAAGCTCTT GCTGAAAACC CTGCCCTTGAT AGAAGGGATG
 451 AAAAGATGC AAGGCCGTGA TTGGATTTGG AATCTTTCT TAACACAATT
 501 AAGTGAAGTA TTTTCTCAAG CTTGGTCTCA AGGGGTTATC TCTGAAGAAG
 551 ATATCGCCGC ATTTGCCCTCC ACCTTAGGTT TGGACTCCGG GACCGTTGCG
 601 TCCATTGTCC AAGGGGAAAG GTGGCCCGAG CTTGTGGATA TAGTGATAAC
 651 TAA

The PSORT algorithm predicts periplasmic (0.925).

This shows that cp7193 is an immunoaccessible protein in the EB and that it is a useful immunogen.
 40 These properties are not evident from the protein's sequence alone.

It will be appreciated that the invention has been described by way of example only and that modifications may be made whilst remaining within the spirit and scope of the invention.

CP7342P	GTGCGT CATATG · AAAAAAAAATTATTTCTACT	ACTCGCTA GCGGCCGC CACACTCTGTTCTCTG
CP7347P	GTGCGT CATATG TTTTCTAAGGATTGACTAA	GCGT CTCGAG CGAACAGAAGTCGT
CP7353P	GTGCGT CATATG AATATGCCCTGTTCTCT	GCGT CTCGAG GGGGCCCTAGGTTGTA
CP7193P	GTGCGT CATATG TGTTCCCTGGATCCT	ACTCGCTA GCGGCCGC AGTTATCACTATATCCACAAG
CP7248P	GTGCGT GCTAGC CTTGAACATTCTAAACAAGAT	GCGT CTCGAG ACGTACTTTAAAGAGCAGACT
CP7261P	GTGCGT CATATG TGTCTATCTGCCATACAG	GCGT CTCGAG TTTTGATGCTCTCTTC
CP7280P	GTGCGT CATATG GACCAGAAAATTGAAAAA	GCGT CTCGAG AGAGGTCTTCGAGTGC
CP7302P	GTGCGT CATATG AATTTCATTTGAGTAGT	GCGT CTCGAG AAACAGTTCGATTTGTG
CP7306P	GTGCGT CATATG CTTCCATTATCAGGGCA	ACTCGCTA GCGGCCGC TTCTTCAGGTTTCAGG
CP7367P	GTGCGT GCTAGC CGTTATGCCGAGGTC	GCGT CTCGAG TTCTGTCATTTGGTG
CP7408P	GTGCGT CATATG TTGAAAATCCAGAAAAA	GCGT CTCGAG ATTCAATTTCGGAAGAG
CP7409P	GTGCGT CATATG AGACGTTATCTTTCATGGT	GCGT CTCGAG CCCCTTGCTCTTTACATAG
CP6733P	GTGCGT ACTAGT TGTCACTTACAGTCAGTAG	GCGT CTCGAG GAATCGGAGTTGGTA
CP6728P	GTGCGT ACTAGT AAGTCCCTGTCCTCTGG	GCGT CTCGAG GAAACAAAATTAGAGCCC

TABLE III – Proteins with best results in FACS analysis

cp number	Molecular Weight (kDa)		Fusion type
	Theoretical	Western Blot	
6260	97.5	94; 70	GST
6270	87.5	-	GST
6272	78.0	90	GST
6273	58.6	74; 64; 50	GST
6296	31.1	-	GST
6390	88.9	102	GST
6456	42.5	89; 67, 45	GST
6466	57.5	59; 56	His
6467	59.0	67	GST
6552	28.4	50; 27	GST
6576	86.0	79; 70; 62; 45	GST
6577	17.3	12	GST
6602	43.4	53; 42; 34	GST
6664	54.5	104; 45	GST
6696	47.9	95; 53	GST
6727	130.0-142.9	123; 61; 39	His
6729	94.8	multiple bands	GST
6731	95.5	97	GST
6733	97.1	104	His
6736	100.1	98; 93; 66; 60	GST
6737	101.2	multiple bands	GST
6751	100.2	95; 71	GST
6752	102.1	97; 48	His
6767	29.1	28	GST
6784	32.9	35	GST
6790	71.3	multiple bands	His
6802	29.7	-	GST
6814	29.6	28	GST

CP6607P	GTGCGT	CATATG	CCTCGTGGTGACACTTT	GCGT	CTCGAG	CGCTGCTTCCTGCTC
CP6615P	GTGCGT	CATATG	TGCTCTCAAAAACGACAA	GCGT	CTCGAG	TGAAGAGGCCGCATC
CP6624P	GTGCGT	CATATG	GATGCGAAAATGGGA	GCGT	CTCGAG	TCTTTGACATTCAAGAGC
CP6672P	GTGCGT	CATATG	ATTCCCTACCATGTTAATG	GCGT	CTCGAG	GTCATACAATTCTTATATA
CP6679P	GTGCGT	CATATG	TGCACTCACTTAGGCT	GCGT	CTCGAG	CGAGTAGTTAGGCACAAAC
CP6717P	GTGCGT	GCTAGC	AAGACAATCGTAGCTTC	ACTCGCTA	GCGGCCGC	GGCTGGCATATAGGT
CP6784P	GTGCGT	GCTAGC	AAATCAAGATGTTCTATTGATA	GCGT	CTCGAG	TCCAAAACAACCCCTCT
CP6802P	GTGCGT	CATATG	TGCGTAAGTTATTAATTCTT	GCGT	CTCGAG	CAGTCGGGCTTGTG
CP6847P	GTGCGT	CATATG	TGGATCTTTACGAG	GCGT	CTCGAG	TTTCTACACTGTTGTAATAAA
CP6884P	GTGCGT	CATATG	AATCAGCTGCTTCT	GCGT	CTCGAG	AGAGAAGGTAATTGTACC
CP6886P	GTGCGT	CATATG	TGTCTACTTATTATCTATCTAC	GCGT	CTCGAG	TTCAGAAAAATGGCT
CP6890P	GTGCGT	CATATG	TCCCCACGACGACAA	GCGT	CTCGAG	TCTTGACGATTTAGC
CP6960P	GTGCGT	CATATG	TGTGACGTACGGTCTA	ACTCGCTA	GCGGCCGC	TTCACCTTGATTTCT
CP6968P	GTGCGT	CATATG	TGGATGCAAAC	ACTCGCTA	GCGGCCGC	GGAAGTATGCTTAGATATT
CP6969P	GTGCGT	CATATG	TGCTGTGGTACTCTATT	ACTCGCTA	GCGGCCGC	AAAAGGTCATAGTATACCT
CP7005P	GTGCGT	CATATG	AAAATGTGATATTGAACA	GCGT	CTCGAG	CTGAGCTCTATTCTATTAT
CP7072P	GTGCGT	CATATG	CCCATTTATGGAAA	GCGT	CTCGAG	GTTGAGCAAGGTTG
CP7101P	GTGCGT	CATATG	TATTCGTGTACAGAA	GCGT	CTCGAG	GAAAAATTCTTAGGGAG
CP7102P	GTGCGT	CATATG	GCCGCTAAAGCAAAT	GCGT	CTCGAG	TGAAAATGAAAGATGGT
CP7105P	GTGCGT	GCTAGC	AGTCTATATCAAAATGGTG	GCGT	CTCGAG	ATCTTCATTTGGTTATCT
CP7106P	GTGCGT	CATATG	AAAGATTTGGGGACTCT	GCGT	CTCGAG	GAATCCTAAGGCATACCTA
CP7107P	GTGCGT	GCTAGC	AGTATAGTCAGAAATTCTCA	GCGT	CTCGAG	GAAGCTAAGATTATAGCTACTTT
CP7108P	GTGCGT	GCTAGC	GCGGCCCTTCCA	ACTCGCTA	GCGGCCGC	TTTATGTATATGAAACAGATAGG
CP7109P	GTGCGT	CATATG	GGACATTTTATTGATATTG	ACTCGCTA	GCGGCCGC	ATCATCAAGGTAGATAAAG
CP7110P	GTGCGT	CATATG	GGTATTGCTATGTAATTACA	GCGT	CTCGAG	TTCTGATTGGACTCCA
CP7127P	GTGCGT	CATATG	GTGGCTTTAACGATAGC	ACTCGCTA	GCGGCCGC	GCAGCCATCGTATTC
CP7130P	GTGCGT	CATATG	TTCAATATGCGAGG	GCGT	CTCGAG	CTTCTTATTGAACTTIG
CP7140P	GTGCGT	CATATG	ACAGCCGGAGCAGCT	GCGT	CTCGAG	AGCACCCCTAATTTCATTG
CP7182P	GTGCGT	CATATG	GGATATGTTCTATGTGATC	GCGT	CTCGAG	GCTACTAAATCGAATCGA
CP6262P	GTGCGT	CATATG	ATCCCCTGATTAAGTCA	ACTCGCTA	GCGGCCGC	TTCACTGGGAGCTTGA
CP6269P	GTGCGT	CATATG	TACCAAGGAAATCTAAGAT	ACTCGCTA	GCGGCCGC	GATTTCTCTCTCAGCTC
CP6296P	GTGCGT	CATATG	GAGGAGGTGCTGAGTAT	ACTCGCTA	GCGGCCGC	ATGTTCTTTACTCTTTCT
CP6419P	GTGCGT	CATATG	GCTCCAGTCCTGTGTT	GCGT	CTCGAG	AAAGTGTCTGGAGAAGT
CP6601P	GTGCGT	CATATG	AATAAGCTACTCAATTCTG	GCGT	CTCGAG	GAAAATCTGAATTCTCT
CP6639P	GTGCGT	CATATG	TTAAATTCAAGCAATTCA	GCGT	CTCGAG	AGGAACCTAAACCTCATCT
CP6664P	GTGCGT	GCTAGC	GTTTTATTCATGCTAA	ACTCGCTA	GCGGCCGC	CTTAAAGAGCTATTTCTAAGTA
CP6696P	GTGCGT	CATATG	TGCGTGATAATGGG	GCGT	CTCGAG	ATTCATCTTGTAAAGAAT
CP6757P	GTGCGT	CATATG	GCAGTTGGTGGCGT	ACTCGCTA	GCGGCCGC	CTGTCCTCTGGAGC
CP6790P	GTGCGT	GCTAGC	AGTGAAACAAAAAAATCA	ACTCGCTA	GCGGCCGC	CTTATGTCGTTATCAATA
CP6814P	GTGCGT	CATATG	CATGACGCACCTCTAAG	GCGT	CTCGAG	TACAGCTGGCGA
CP6834P	GTGCGT	CATATG	GTTATGGGAACTTATATCG	GCGT	CTCGAG	TACATTGTTGATTTCAAG
CP6878P	GTGCGT	CATATG	AACGCCCTGATTC	GCGT	CTCGAG	GCTAGCGCTTTC
CP6892P	GTGCGT	CATATG	CAGAACATCTTCT	ACTCGCTA	GCGGCCGC	TCCTCTTAGGAAATGG
CP6909P	GTGCGT	CATATG	TCCTCTTAGGAAATGG	GCGT	CTCGAG	CAGTGCCAAGTAGGGA
CP7015P	GTGCGT	CATATG	GCAGTACGATTAATTGTTG	GCGT	CTCGAG	TTTATTGTTGCTATTTATATTC
CP7035P	GTGCGT	GCTAGC	AGCAGAAAAGACAATGA	GCGT	CTCGAG	ATTTGAGTGTCTTGCA
CP7073P	GTGCGT	CATATG	ATTACCATAAAATCACGTG	GCGT	CTCGAG	TATCCATCGACTTATAGC
CP7085P	GTGCGT	GCTAGC	TGTATTTCCCTACGTA	ACTCGCTA	GCGGCCGC	GGATTCCTGCATACTCTG
CP7092P	GTGCGT	CATATG	TCTCTCTTCTAA	GCGT	CTCGAG	GGATTCAATTACTGACCA
CP7093P	GTGCGT	CATATG	AAATACCGCTTACG	GCGT	CTCGAG	ATTCTGAGGGCTACGT
CP7094P	GTGCGT	CATATG	GTACACTTCTCTACAAACCC	GCGT	CTCGAG	TAAGTTGTTGATTCGGTAT
CP7132P	GTGCGT	CATATG	TTGTTATTAGGGACTTTAGGA	GCGT	CTCGAG	TTTCCCAACCGCA
CP7133P	GTGCGT	CATATG	GCTGCGAATGTC	GCGT	CTCGAG	TAATTAAATACTCTTGAAGG
CP7177P	GTGCGT	CATATG	CCTACTCAAGTAAACAGA	GCGT	CTCGAG	AAGTTTATTTACGCACTT
CP7184P	GTGCGT	GCTAGC	CATATAGGATTTGCCA	GCGT	CTCGAG	GTACTTAGCAAGCGAT
CP7206P	GTGCGT	GCTAGC	AAGAAGCTATATCACCTA	GCGT	CTCGAG	CACACCGAGGAAAC
CP7222P	GTGCGT	CATATG	GTAGTTCAAGAAAAAGTC	GCGT	CTCGAG	ACGTATGCCCAACTG
CP7223P	GTGCGT	CATATG	GAAGTATTAGACCGCTCT	GCGT	CTCGAG	CGAGAAAAAGCTTCC
CP7224P	GTGCGT	CATATG	ATGAAGAAAATCGAAA	ACTCGCTA	GCGGCCGC	TAAGCAATTCAAAATG
CP7225P	GTGCGT	CATATG	CATATTGCTGATCGT	GCGT	CTCGAG	TCTTTAACTAAATCTGTTCTT
CP7303P	GTGCGT	CATATG	CTTGCTTATGTTGATCC	GCGT	CTCGAG	AAAATATACGGAACCTGC
CP7304P	GTGCGT	GCTAGC	GAAGTTATGTTTCCC	GCGT	CTCGAG	TTTTGATTCTTAAGAAG
CP7305P	GTGCGT	CATATG	GAAGTTATGTTTCAACCT	GCGT	CTCGAG	ACTCCCTGAGAAGGGAA
CP7307P	GTGCGT	CATATG	CTTAATCATGCTAAAAGC	ACTCGCTA	GCGGCCGC	CTCTTTATTTAGGAAGCT

CLAIMS

1. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 97, 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 99, 101, 103, 105,
5 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 10 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373, 375, & 377.
2. A protein having 50% or greater sequence identity to a protein according to claim 1.
3. A protein comprising a fragment of an amino acid sequence selected from the group consisting of SEQ IDs 97, 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 20 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373, 375, & 377.
- 25 4. A nucleic acid molecule which encodes a protein according to any one of claims 1 to 3.
5. A nucleic acid molecule according to claim 4, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 98, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 30 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318,

6830	177.4	174; 91; 13	GST
6849	57.3	multiple bands	GST
6850	7.4-9.4	61; 14; 8	GST
6854	42.2	-	GST
6878	40.4	-	GST
6900	28.0	-	GST
6960	25.6	75; 35	GST
6968	34.6	83; 53; 35	GST
6998	39.3	multiple bands	GST
7033	68.2	multiple bands	GST
7101	113	105	GST
7102	63.4	-	GST
7105	29.2	30	GST
7106	39.5	72; 46	GST
7107	71.4	67; 31	His
7108	35.9	35	GST
7111	46.1	51	GST
7132	17.9	57; 47; 17	His
7140	36.2-29.8	50; 38; 34	GST
7170	34.4	77; 33	GST
7224	39.4	40	GST
7287	167.3	180	GST
7306	50.1	50	GST

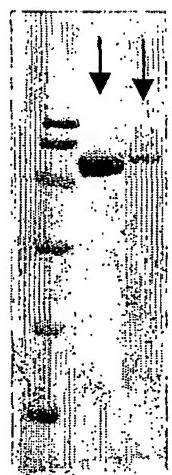
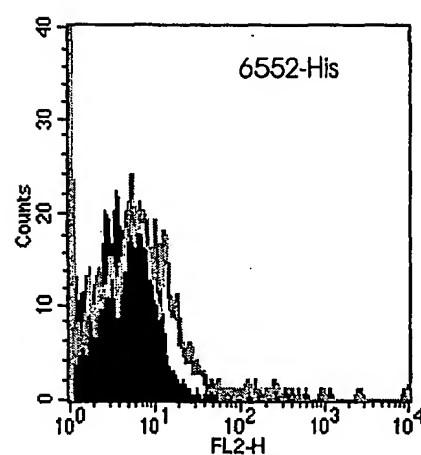
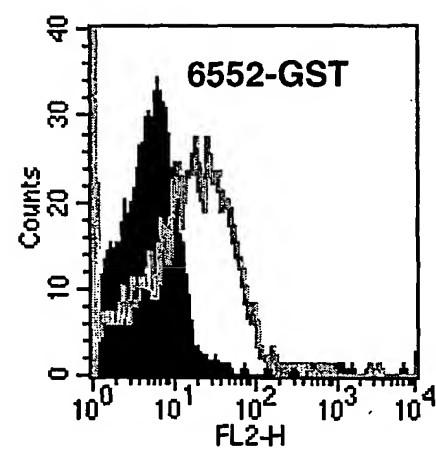
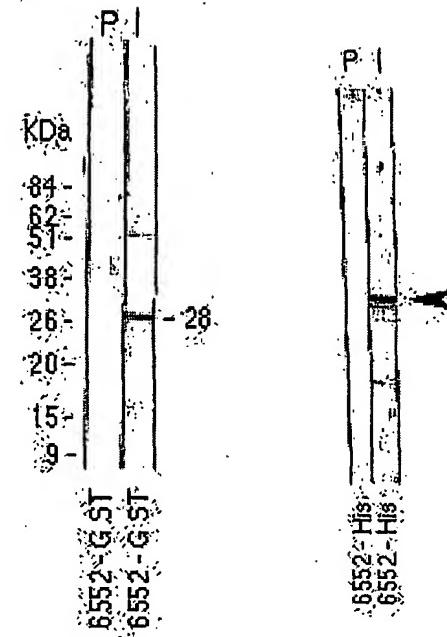
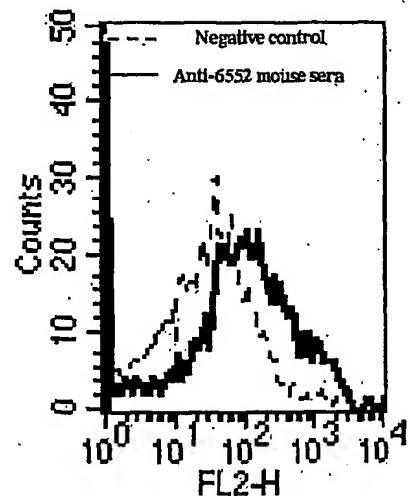
TABLE IV – FACS-positive proteins not found in *C.trachomatis*

cp7105	cp6390
cp7106	cp6784
cp7107	cp6296
cp7108	

TABLE V – Proteins identified by MALDI-TOF following 2D electrophoresis

cp6270	cp6733	cp6900
cp6552	cp6736	cp6960
cp6576	cp6737	cp6998
cp6577	cp6752	cp7033
cp6602	cp6767	cp7108
cp6664	cp6784	cp7111
cp6727	cp6790	cp7170
cp6728	cp6830	cp7287
cp6729	cp6849	cp7306

1/169

FIGURE 1**FIG. 1A****FIG. 1B****FIG. 1C**

320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, & 378.

6. A nucleic acid molecule comprising a fragment of a nucleotide sequence selected from the group consisting of SEQ IDs 98, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, & 378.
7. A nucleic acid molecule comprising a nucleotide sequence complementary to a nucleic acid molecule according to any one of claims 4 to 6.
8. A nucleic acid molecule comprising a nucleotide sequences having 50% or greater sequence identity to a nucleic acid molecule according to any one of claims 4 to 7.
9. A nucleic acid molecule which can hybridise to a nucleic acid molecule according to any one of claims 4 to 8 under high stringency conditions.
- 20 10. A composition comprising a protein or a nucleic acid molecule according to any preceding claim.
11. A composition according to claim 10 being a vaccine composition.
12. A composition according to claim 10 or claim 11 for use as a pharmaceutical.
13. The use of a composition according to claim 10 in the manufacture of a medicament for the treatment or prevention of infection due to *Chlamydia* bacteria, particularly *Chlamydia pneumoniae*.

5

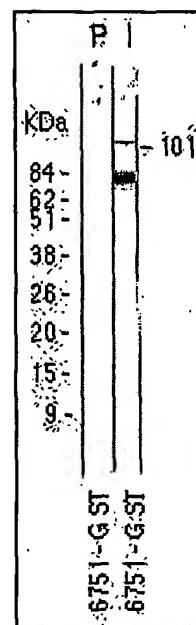
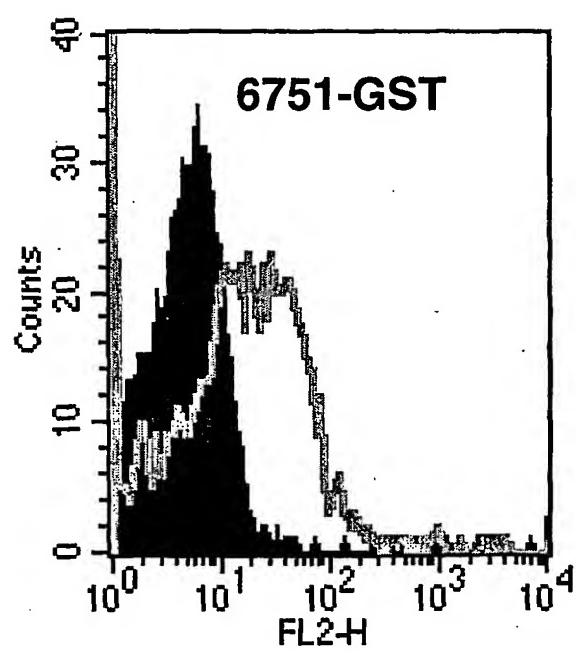
10

15

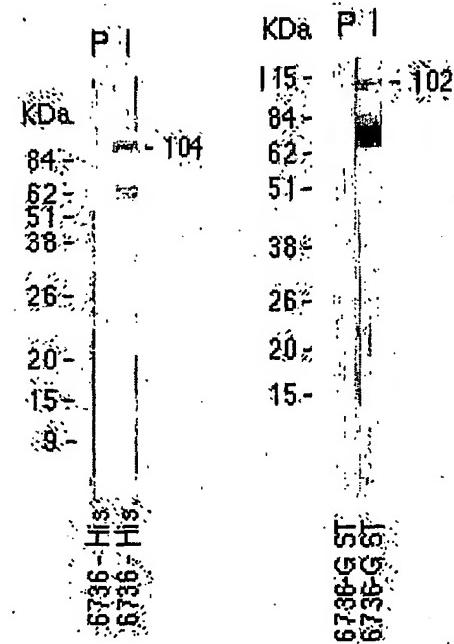
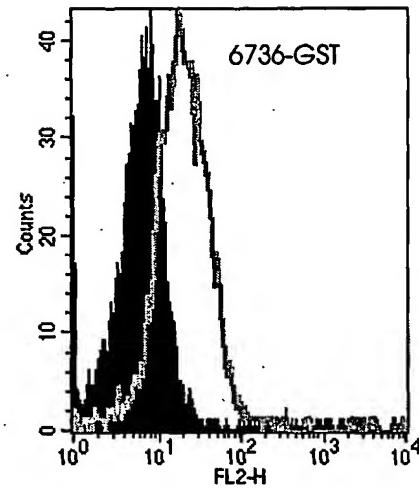
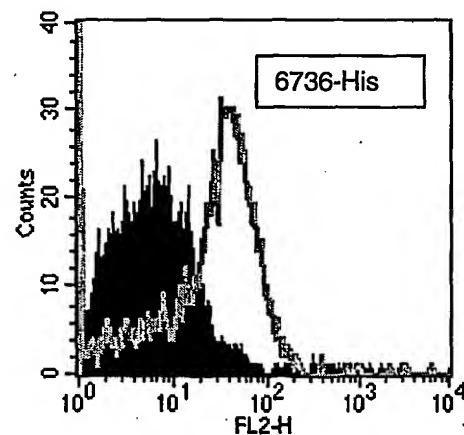
20

25

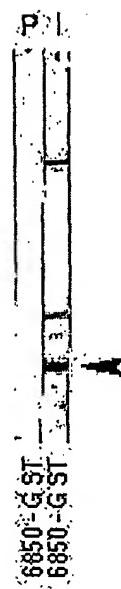
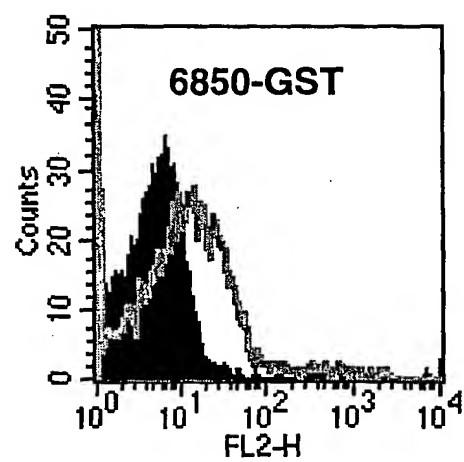
3/169

FIGURE 3**FIG. 3A****FIG. 3B****FIG. 3C**

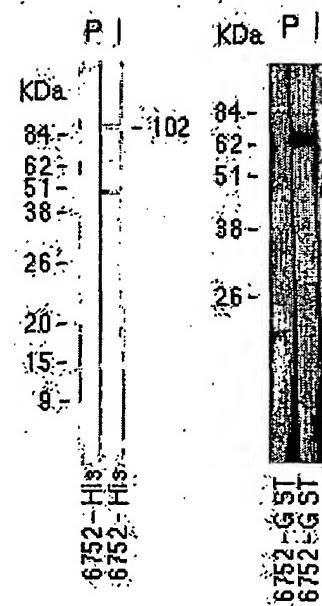
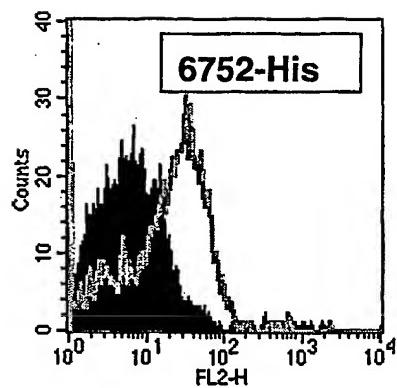
2/169

FIGURE 2**FIG. 2A****FIG. 2B****FIG. 2C**

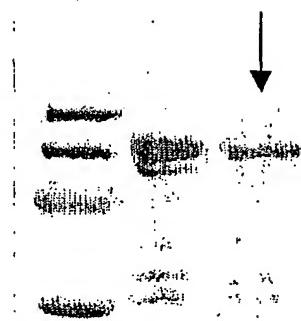
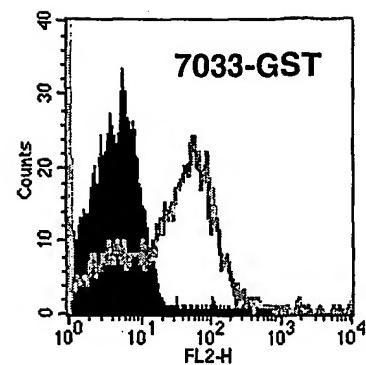
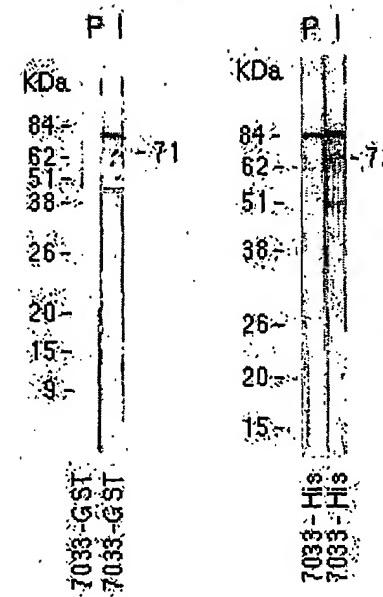
5/169

FIGURE 5**FIG. 5A****FIG. 5B****FIG. 5C**

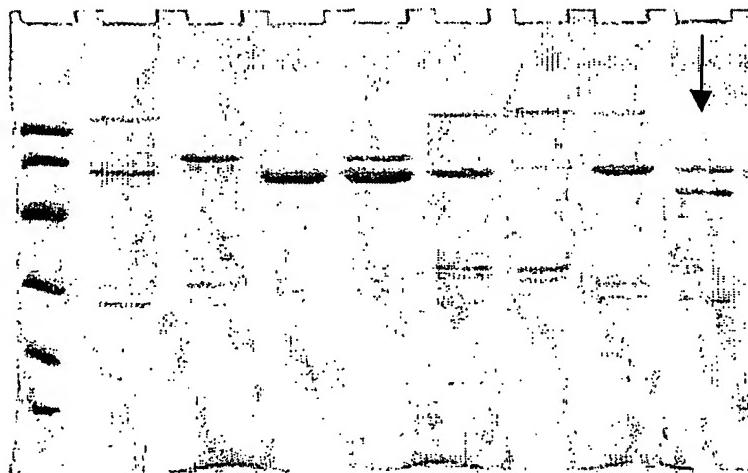
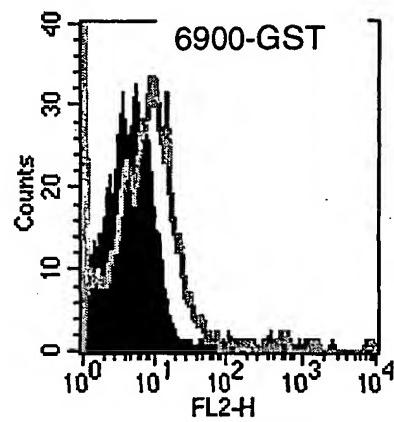
4/169

FIGURE 4**FIG. 4A****FIG. 4B****FIG. 4C**

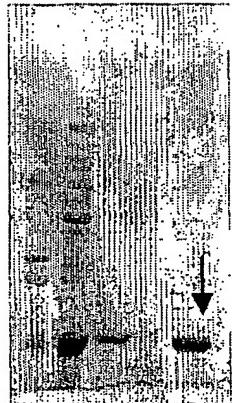
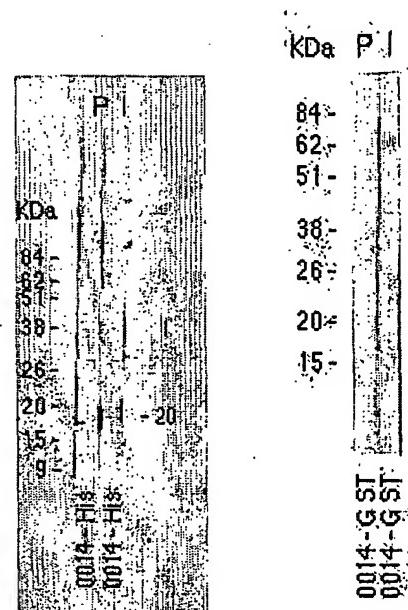
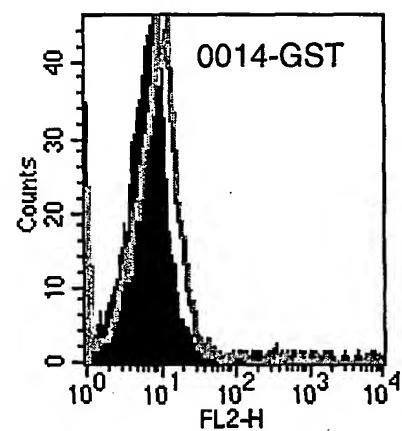
7/169

FIGURE 7**FIG. 7A****FIG. 7B****FIG. 7C**

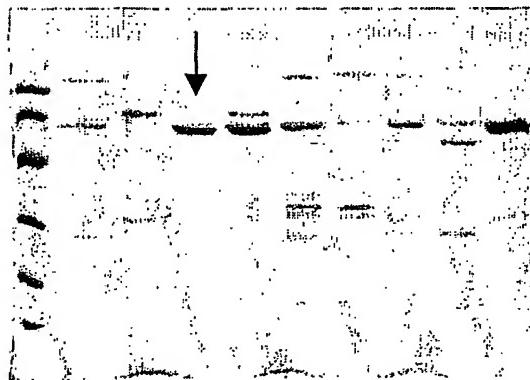
6/169

FIGURE 6**FIG. 6A****FIG. 6B****FIG. 6C**

9/169

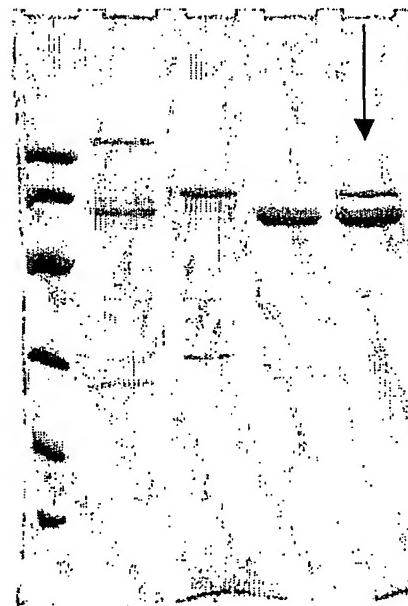
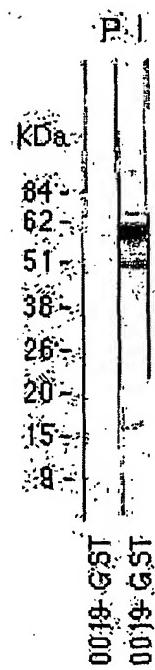
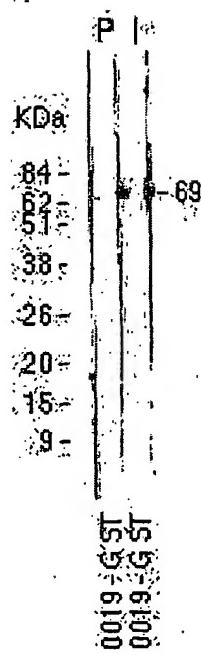
FIGURE 9**FIG. 9A****FIG. 9B****FIG. 9C**

8/169

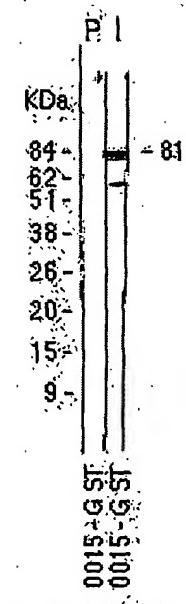
FIGURE 8**FIG. 8A****FIG. 8C****FIG. 8B**

0017-G ST
0017-G ST

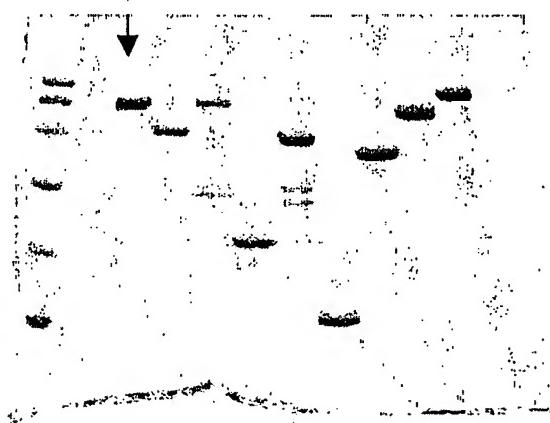
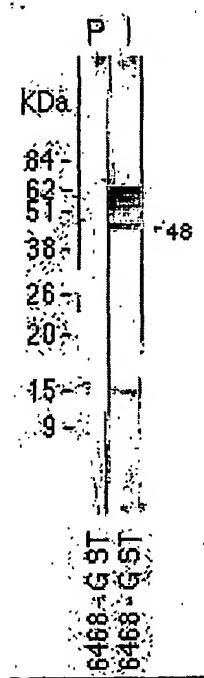
11/169

FIGURE 11**FIG. 11A****FIG. 11B****FIG. 11C**

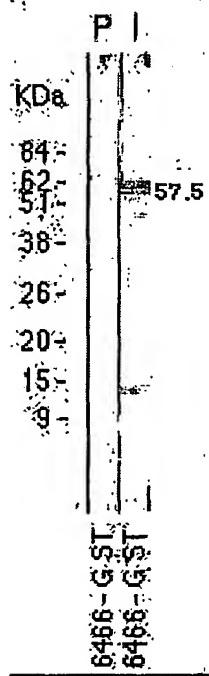
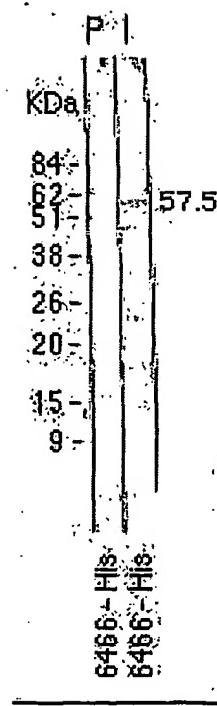
10/169

FIGURE 10**FIG. 10A****FIG. 10B**

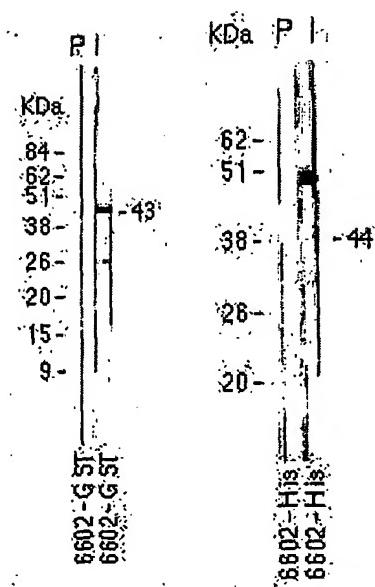
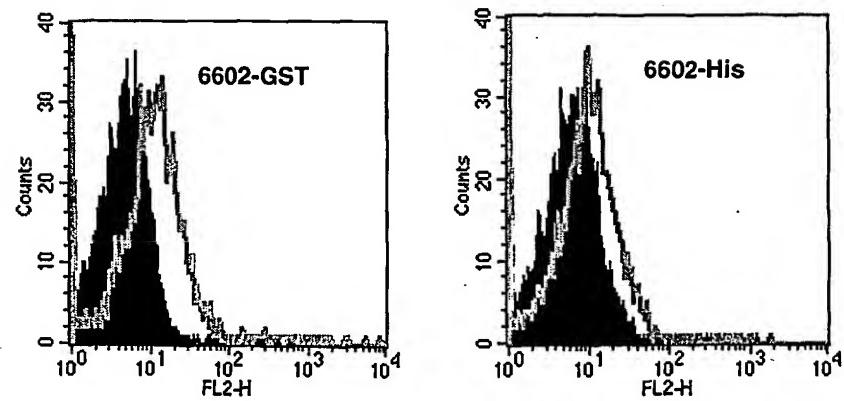
13/169

FIGURE 13**FIG. 13A****FIG. 13B**

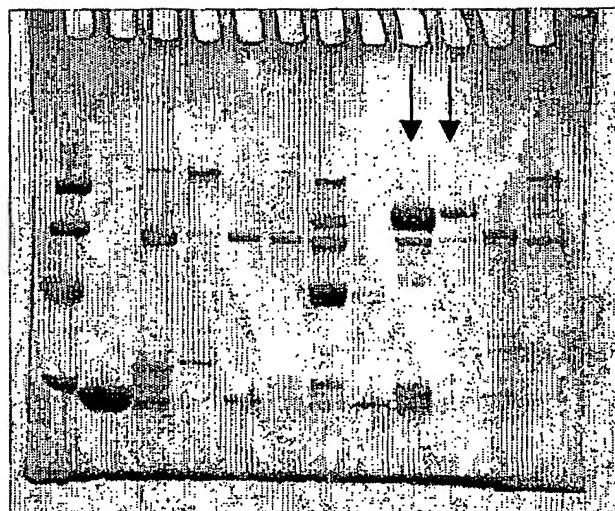
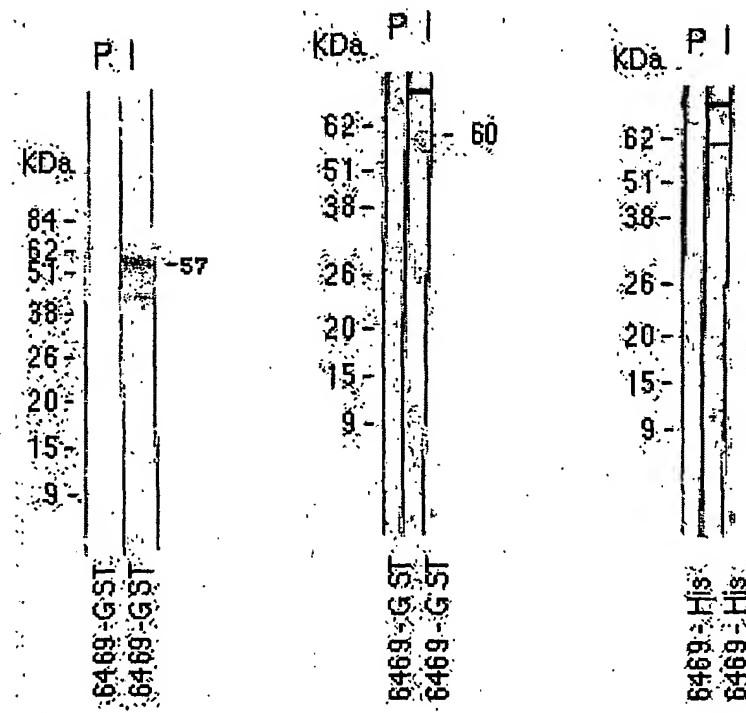
12/169

FIGURE 12**FIG. 12A****FIG. 12B****FIG. 12C**

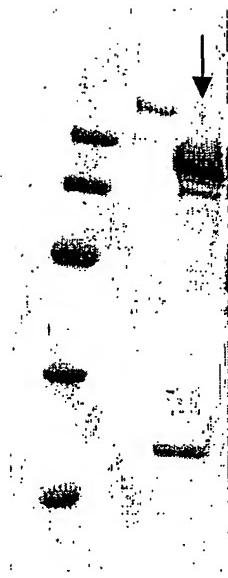
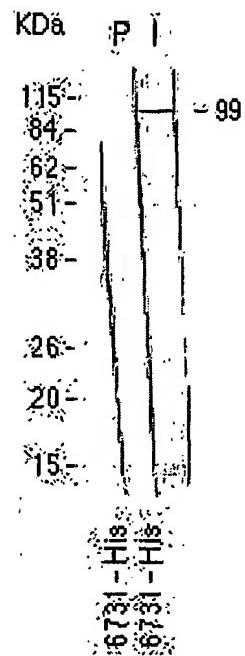
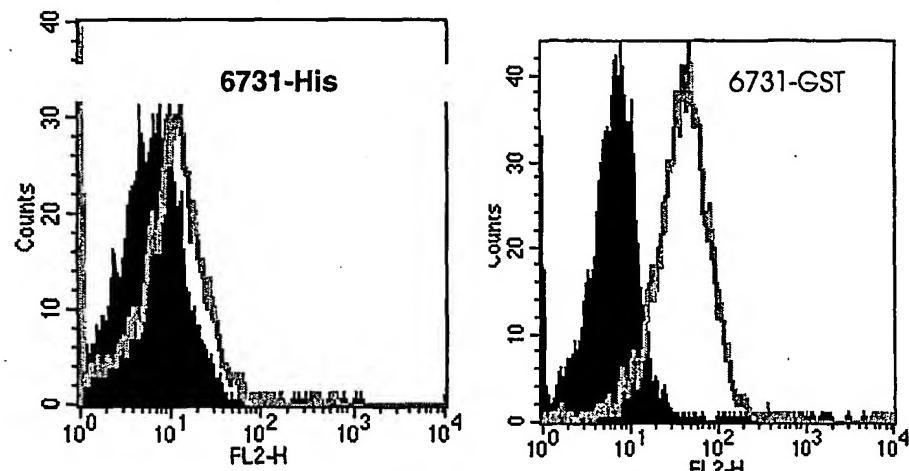
15/169

FIGURE 15**FIG. 15A****FIG. 15B****FIG. 15C**

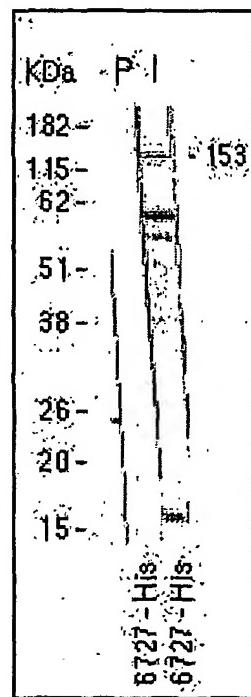
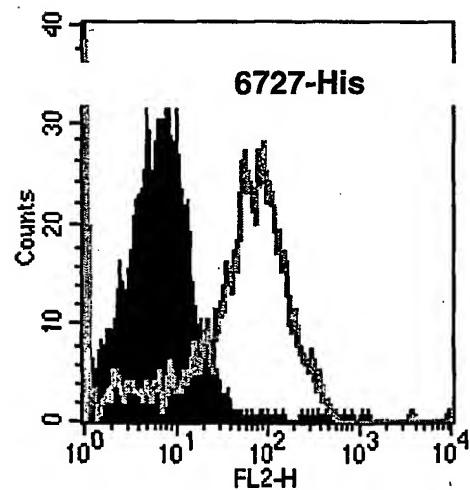
14/169

FIGURE 14**FIG. 14A****FIG. 14B**

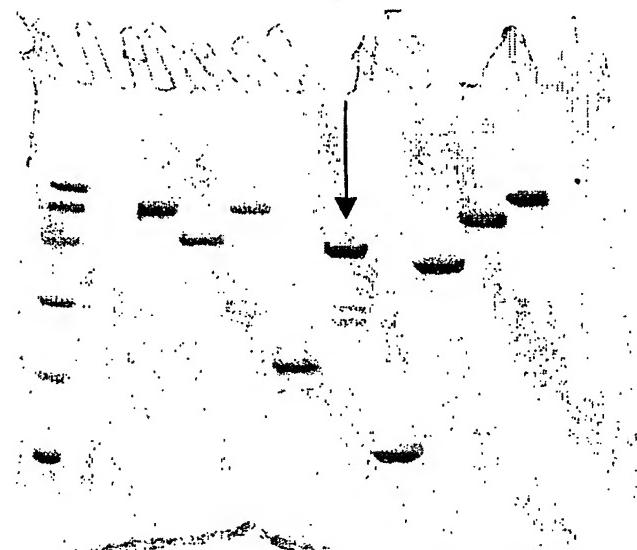
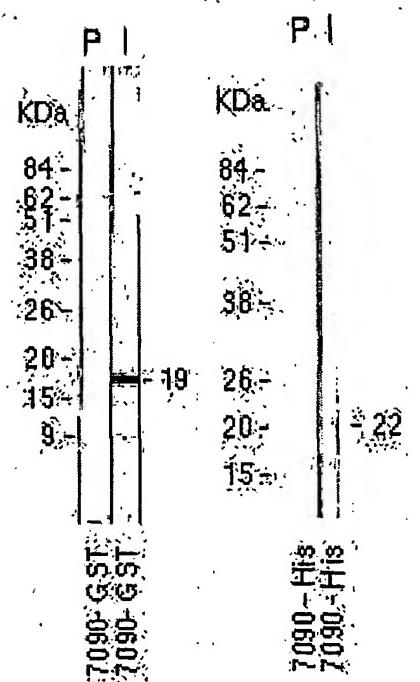
17/169

FIGURE 17**FIG. 17A****FIG. 17B****FIG. 17C**

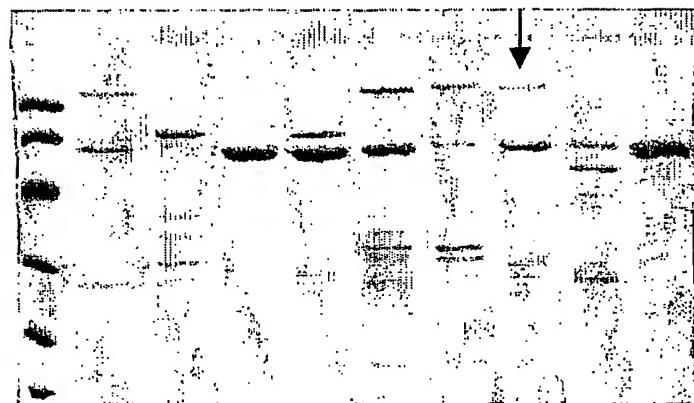
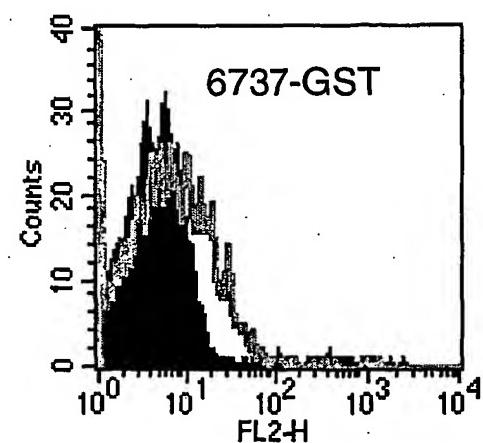
16/169

FIGURE 16**FIG. 16A****FIG. 16B****FIG. 16C**

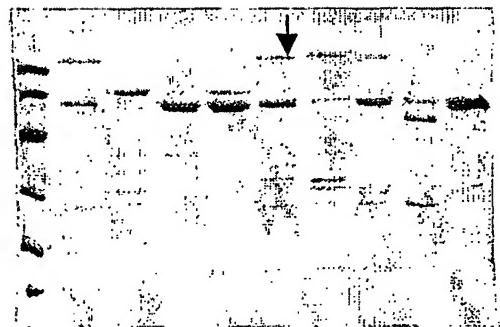
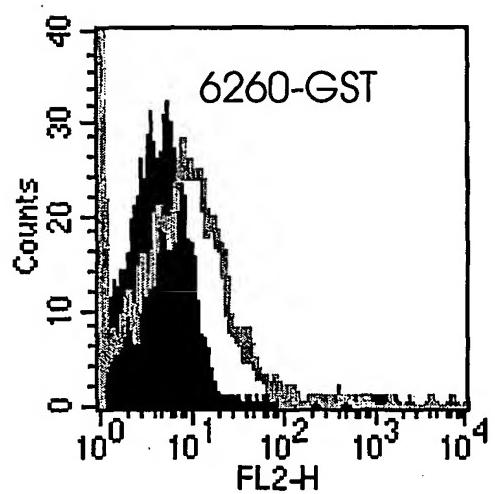
19/169

FIGURE 19**FIG. 19A****FIG. 19B**

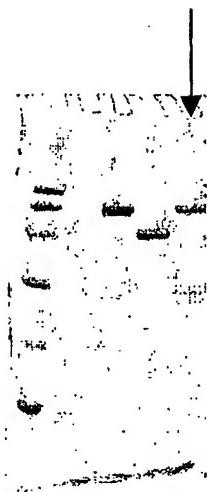
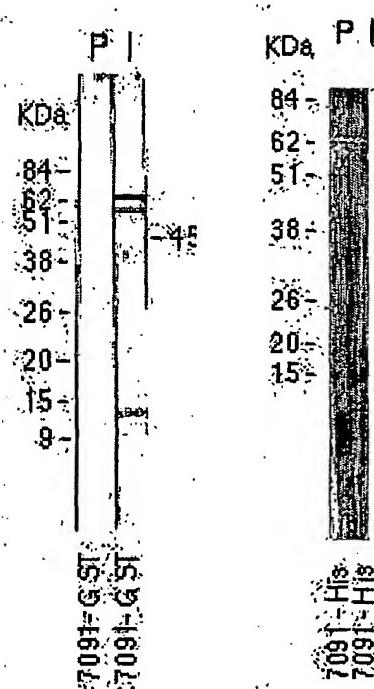
18/169

FIGURE 18**FIG. 18A****FIG. 18B****FIG. 18C**

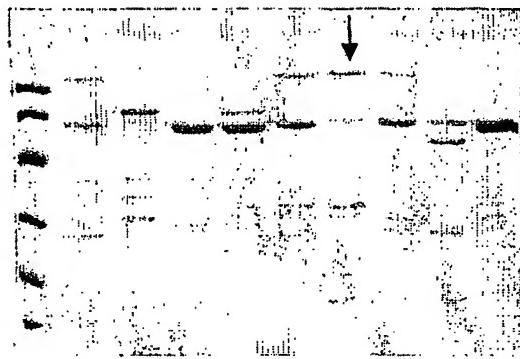
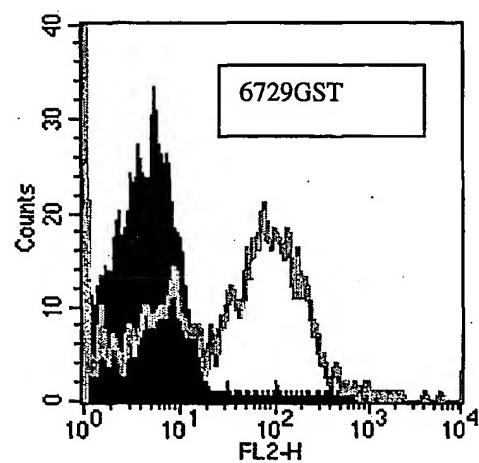
21/169

FIGURE 21**FIG.
21A****FIG.
21B****FIG.
21C**

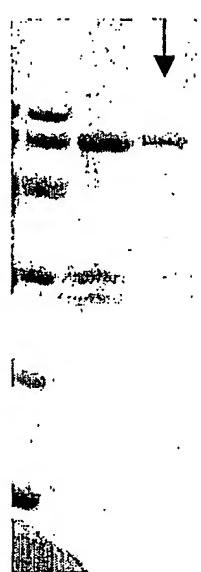
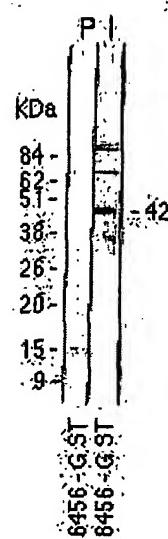
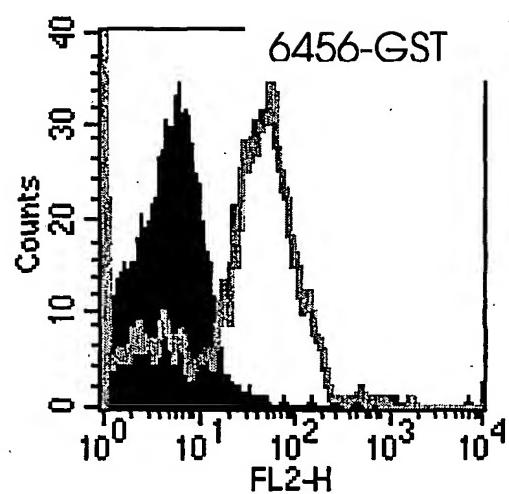
20/169

FIGURE 20**FIG. 20A****FIG. 20B**

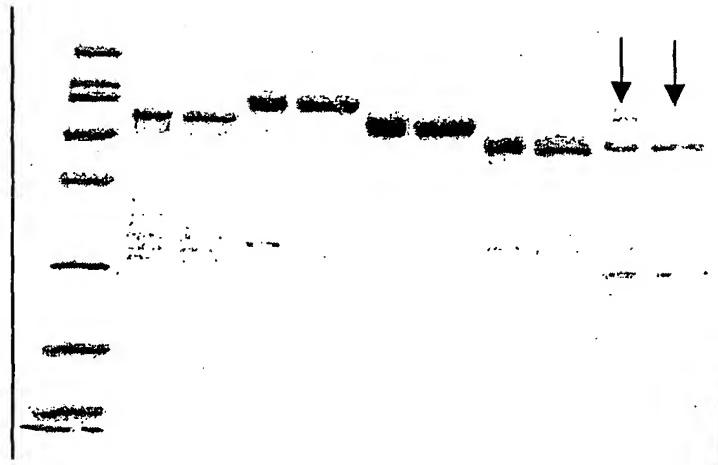
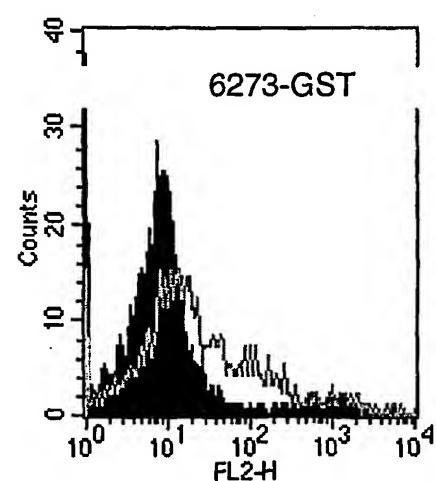
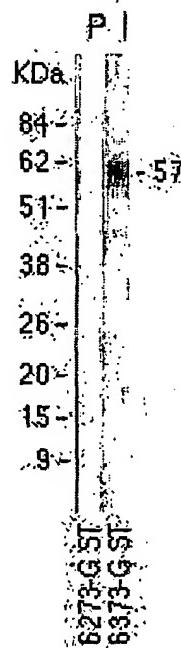
23/169

FIGURE 23**FIG.
23A****FIG.
23B****FIG.
23C**

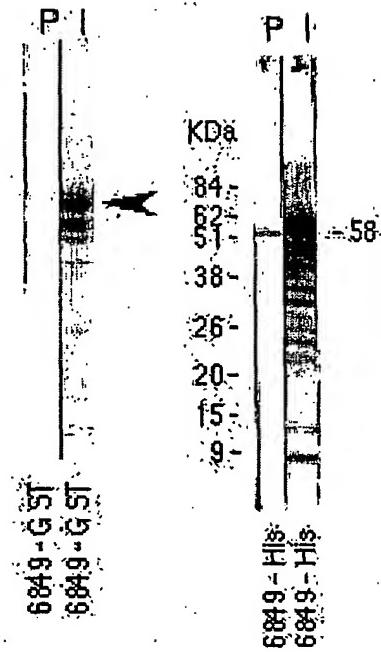
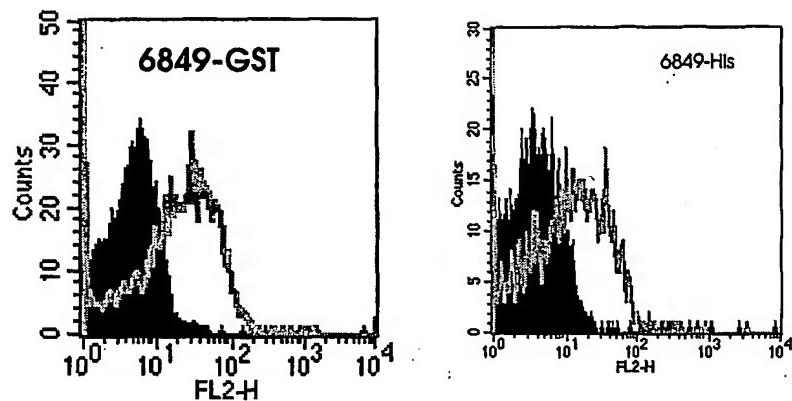
22/169

FIGURE 22**FIG.
22A****FIG.
22B****FIG.
22C**

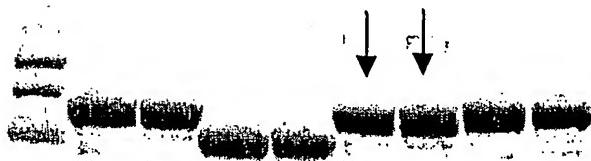
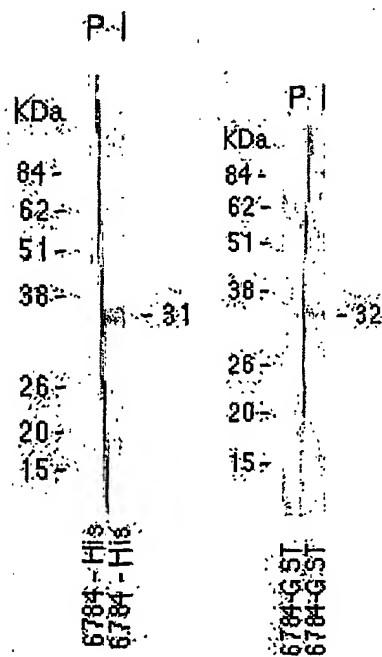
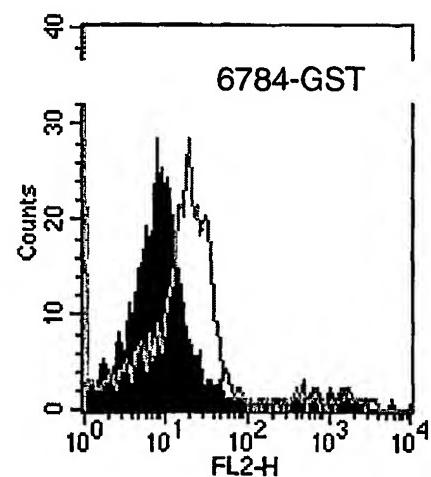
25/169

FIGURE 25**FIG. 25A****FIG. 25C****FIG. 25B**

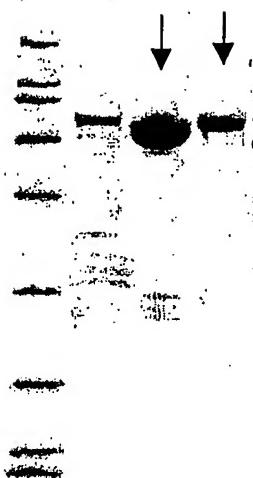
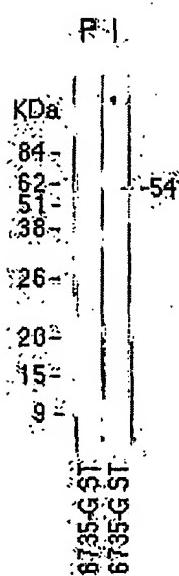
24/169

FIGURE 24**FIG.
24A****FIG.
24B****FIG.
24C**

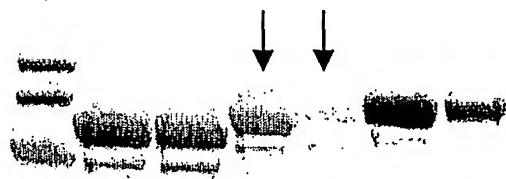
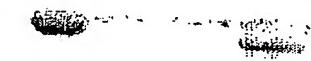
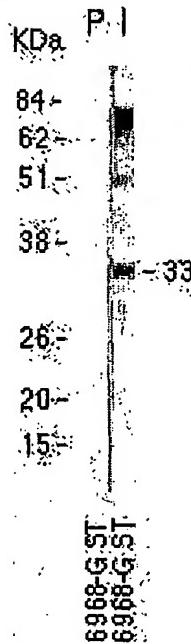
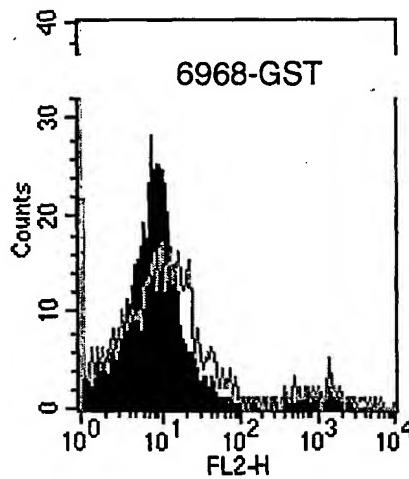
27/169

FIGURE 27**FIG. 27A****FIG. 27B****FIG. 27C**

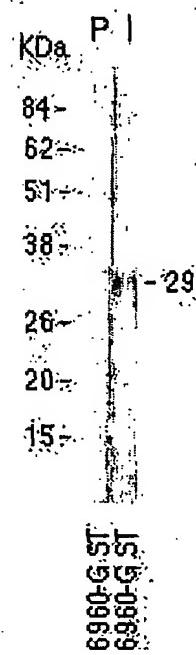
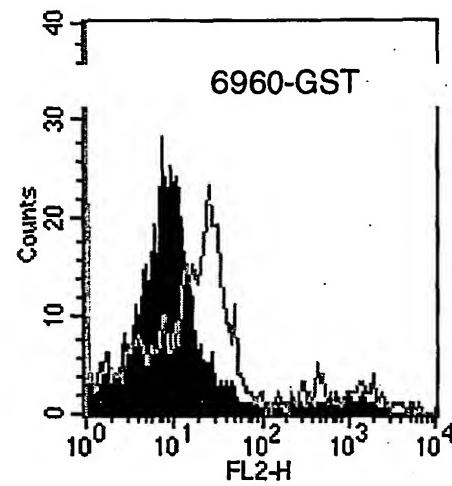
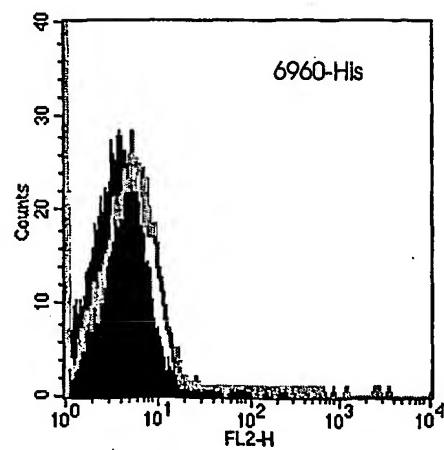
26/169

FIGURE 26**FIG. 26A****FIG. 26B**

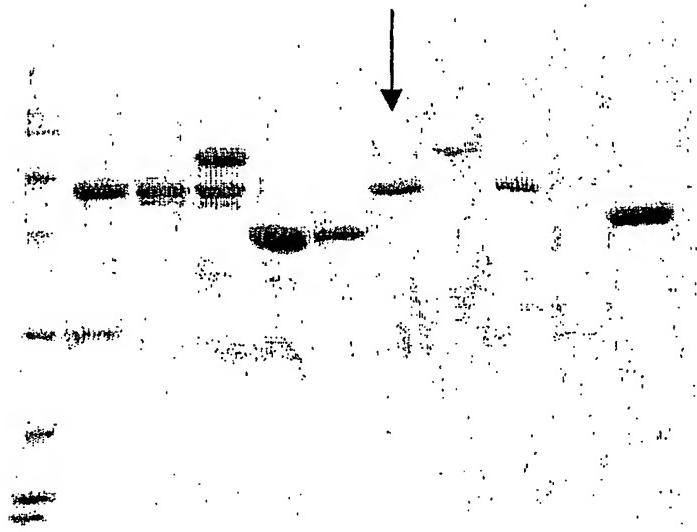
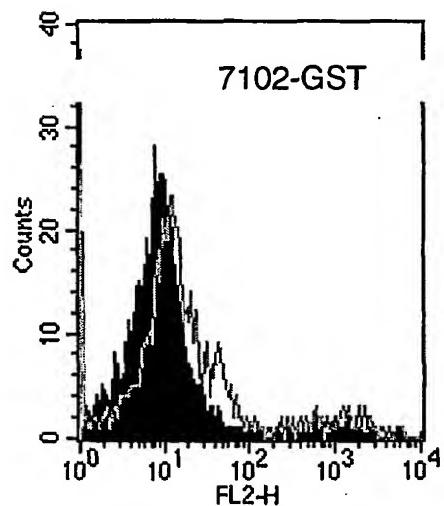
29/169

FIGURE 29**FIG. 29A****FIG. 29B****FIG. 29C**

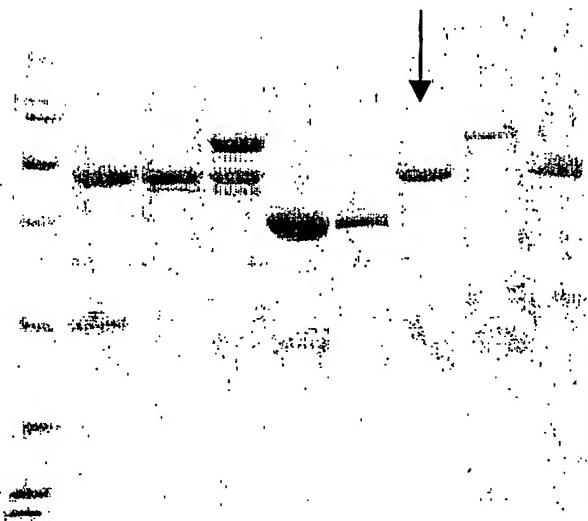
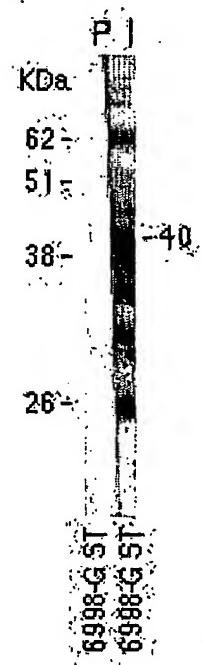
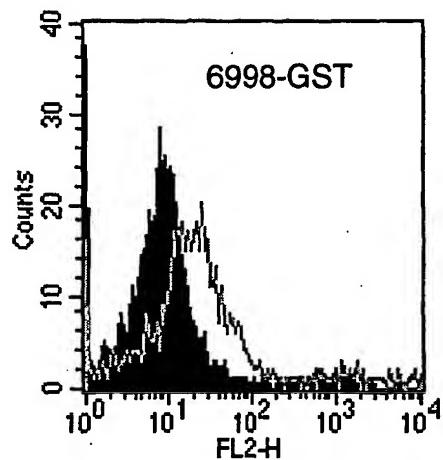
28/169

FIGURE 28**FIG. 28A****FIG. 28B****FIG. 28C**

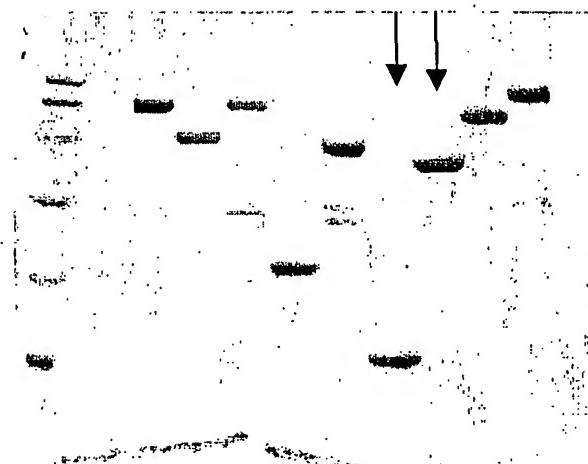
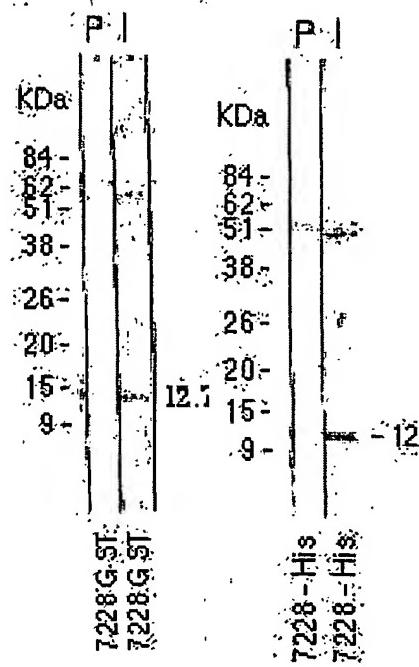
31/169

FIGURE 31**FIG. 31A****FIG. 31B**

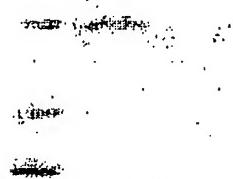
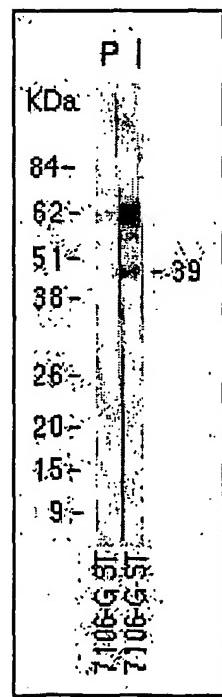
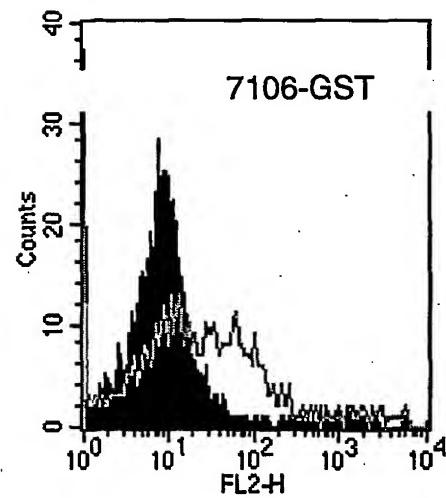
30/169

FIGURE 30**FIG. 30A****FIG. 30B****FIG. 30C**

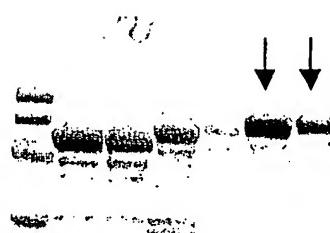
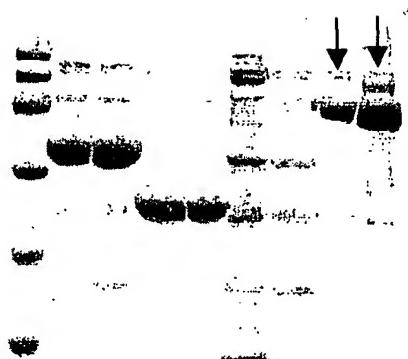
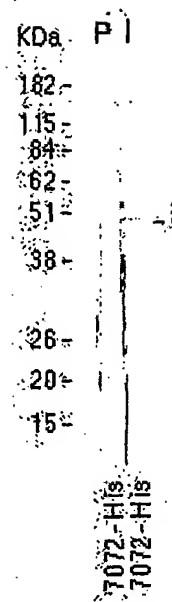
33/169

FIGURE 33**FIG. 33A****FIG. 33B**

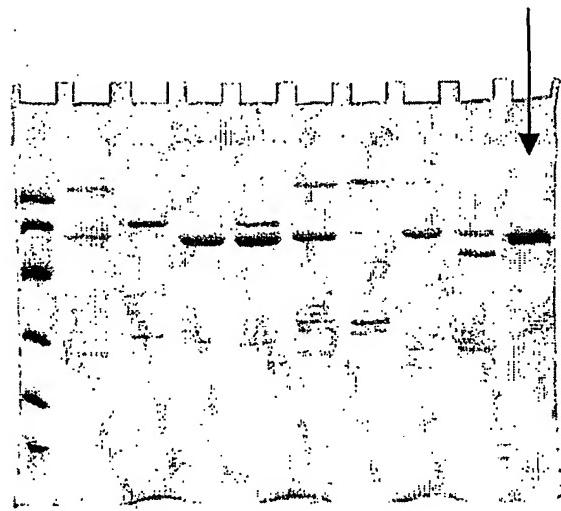
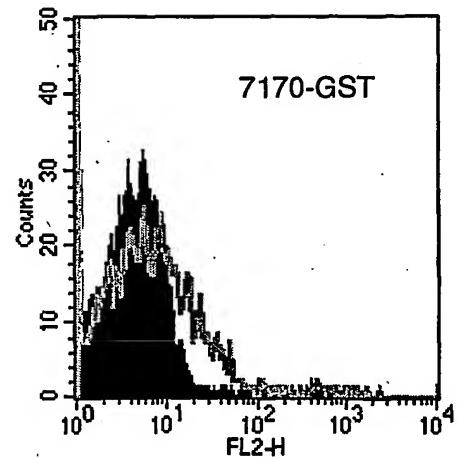
32/169

FIGURE 32**FIG. 32A****FIG. 32C****FIG. 32B**

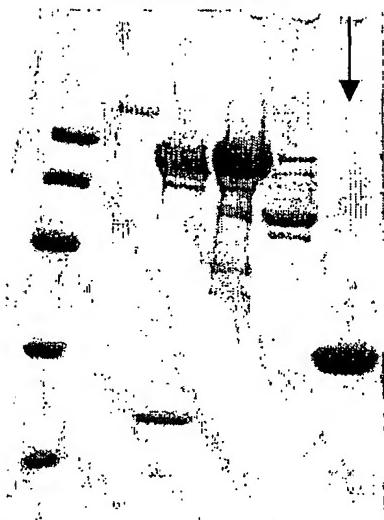
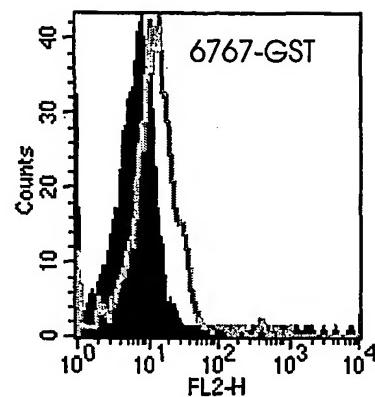
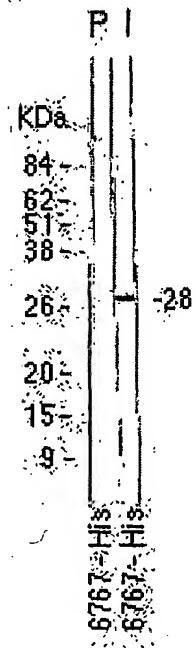
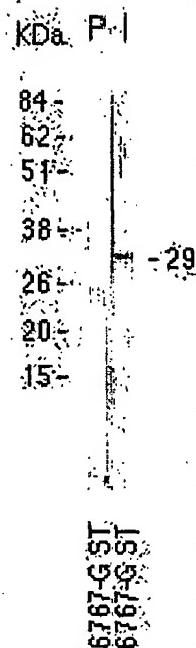
35/169

FIGURE 35**FIG. 35A****FIG. 35B****FIG. 35C**

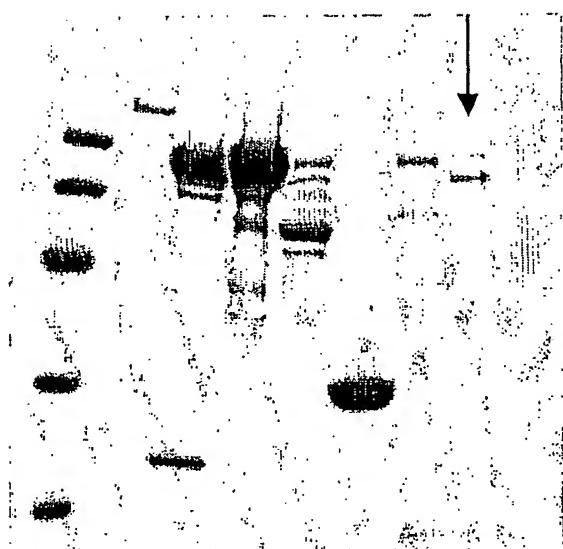
34/169

FIGURE 34**FIG. 34A****FIG. 34B****FIG. 34C**

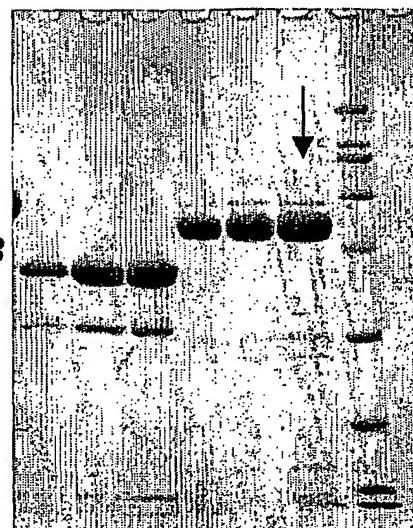
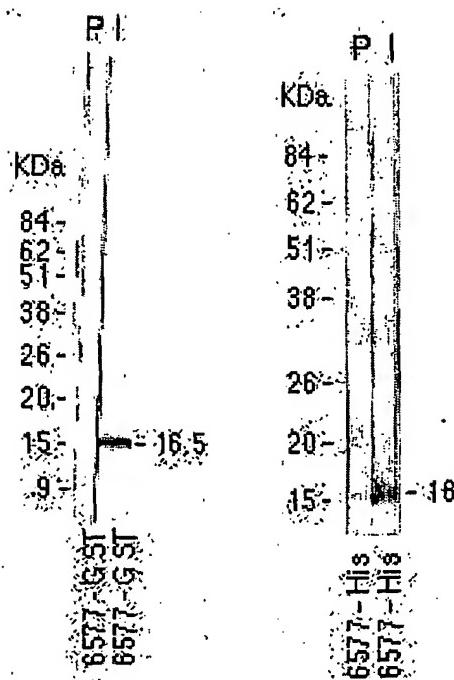
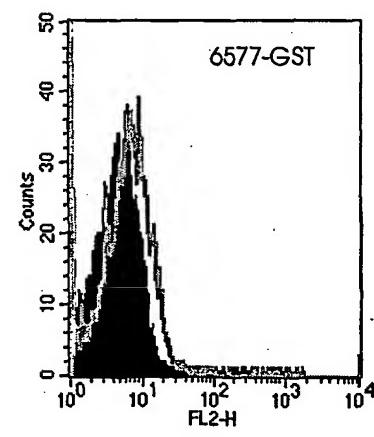
37/169

FIGURE 37**FIG. 37A****FIG. 37C****FIG. 37B****FIG. 37D**

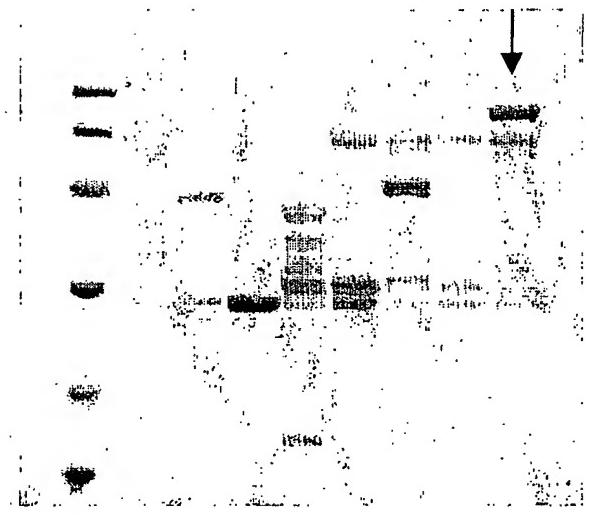
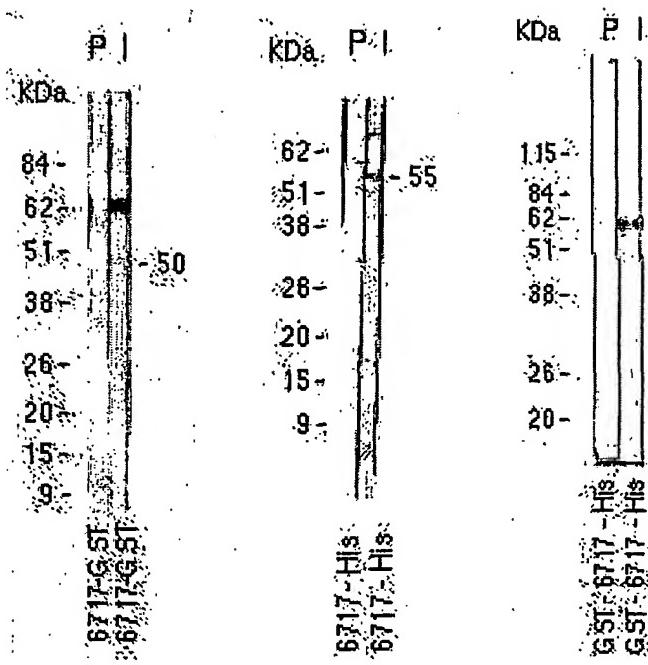
36/169

FIGURE 36

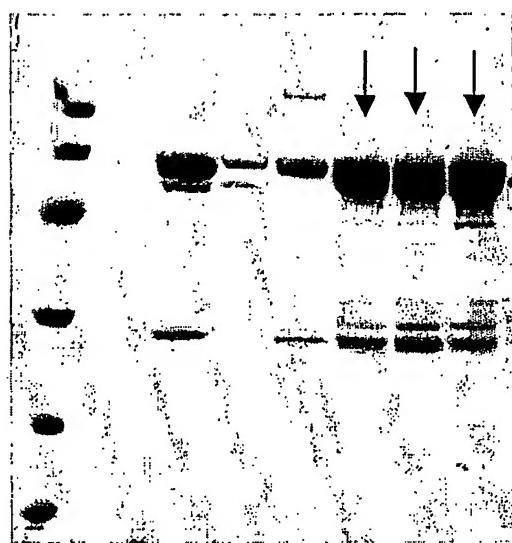
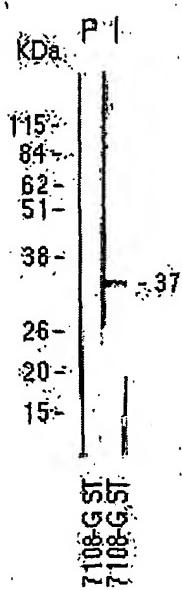
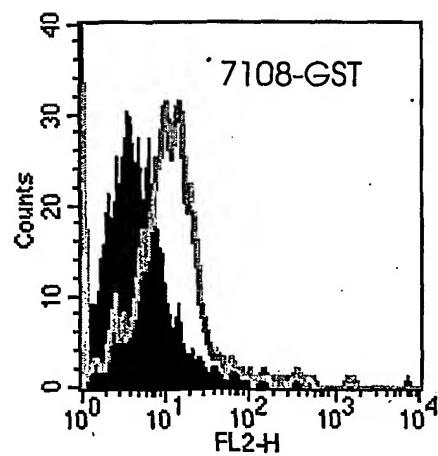
39/169

FIGURE 39**FIG. 39A****FIG. 39B****FIG.
39C****FIG.
39D**

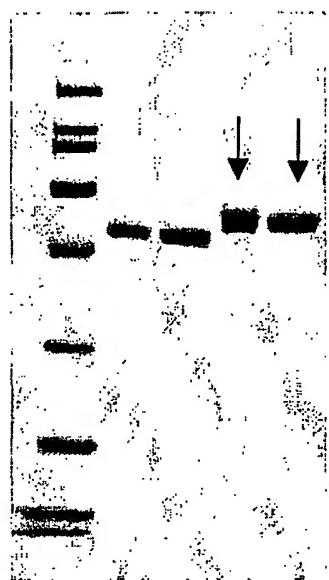
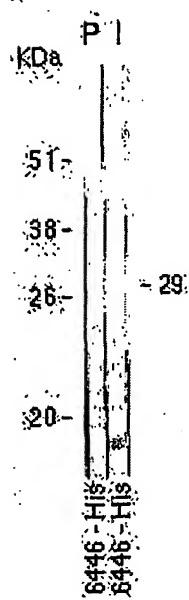
38/169

FIGURE 38**FIG. 38A****FIG. 38B**

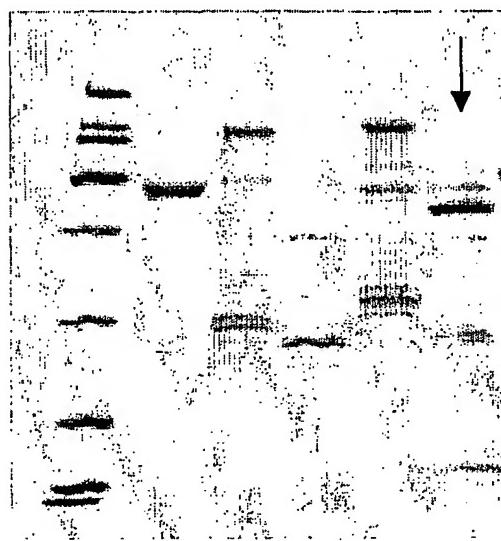
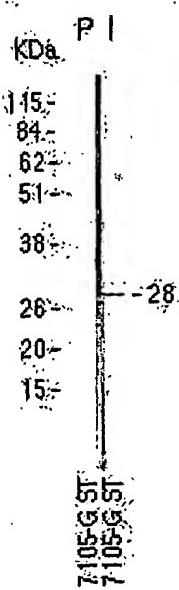
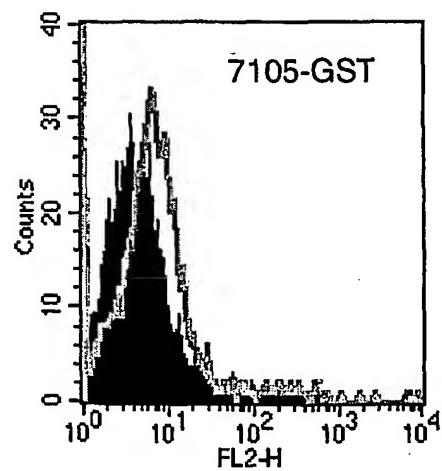
41/169

FIGURE 41**FIG. 41A****FIG. 41B****FIG. 41C**

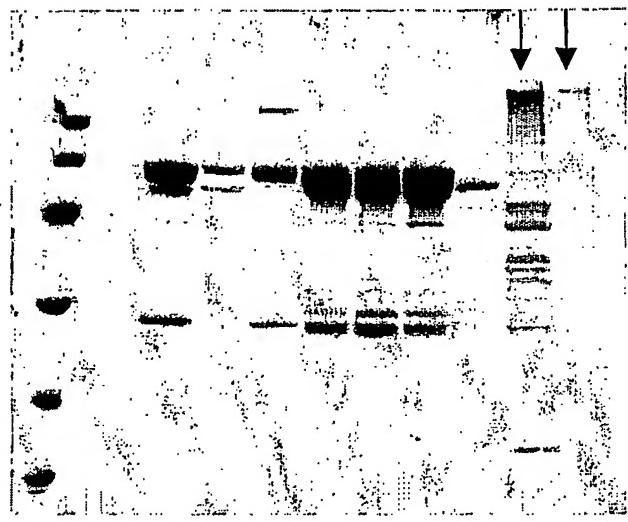
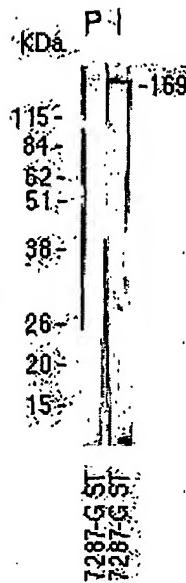
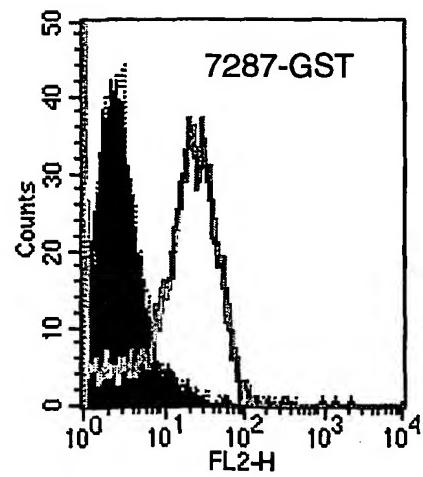
40/169

FIGURE 40**FIG. 40A****FIG. 40B**

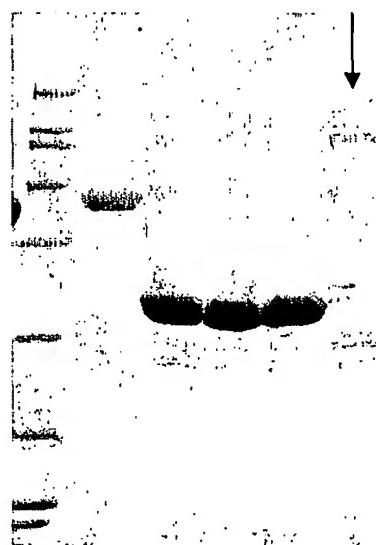
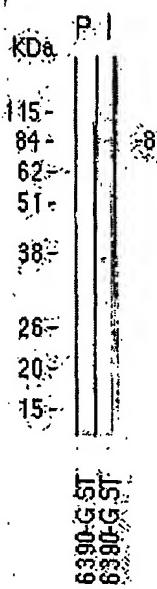
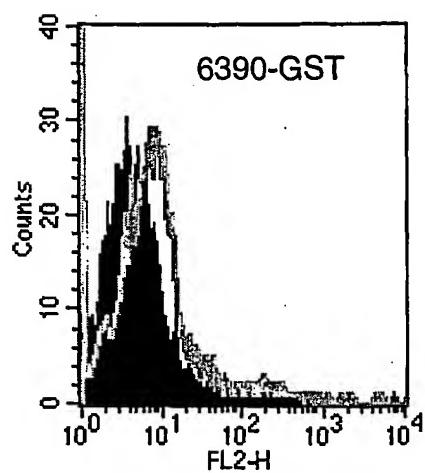
43/169

FIGURE 43**FIG. 43A****FIG. 43B****FIG. 43C**

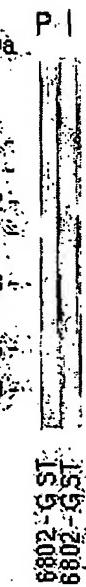
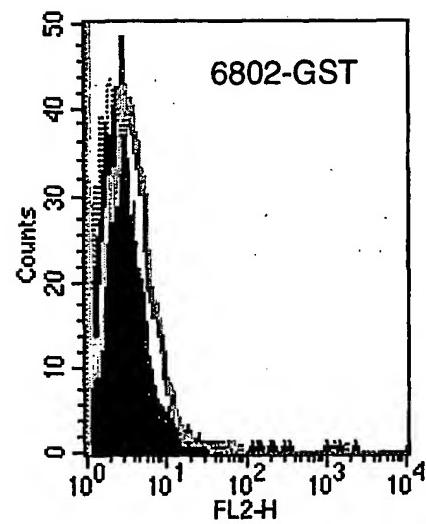
42/169

FIGURE 42**FIG. 42A****FIG. 42B****FIG. 42C**

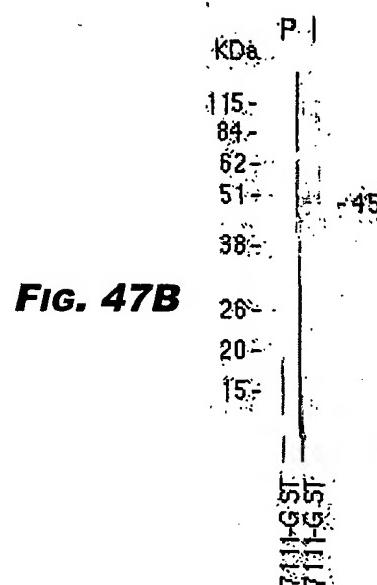
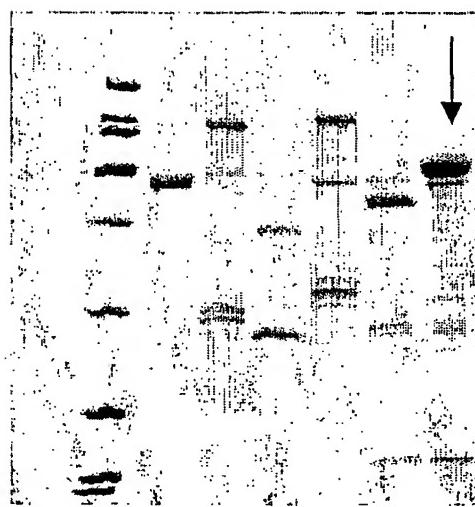
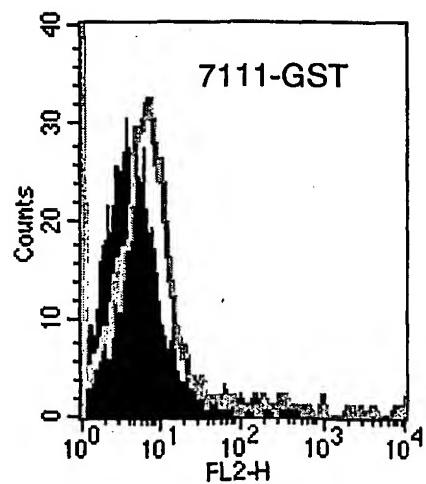
45/169

FIGURE 45**FIG. 45A****FIG. 45B****FIG. 45C**

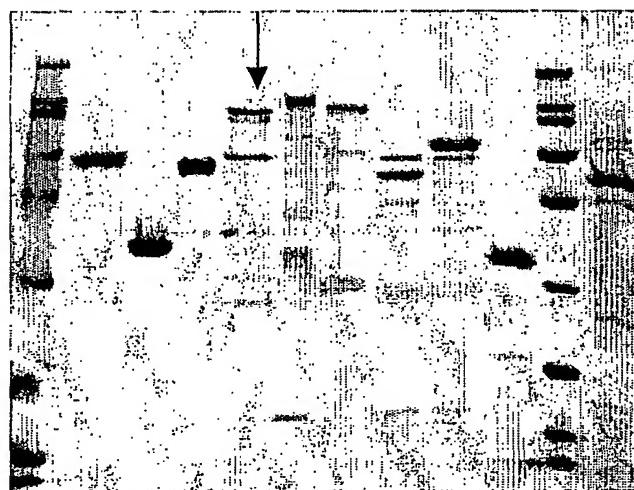
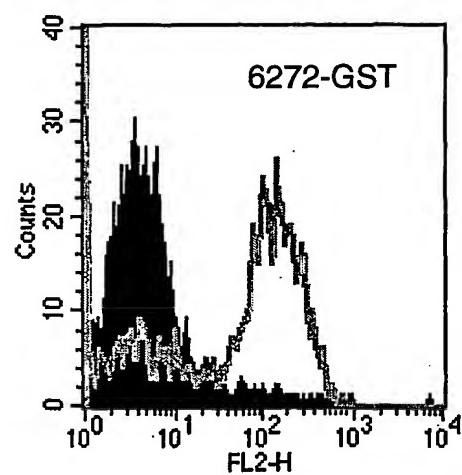
44/169

FIGURE 44**FIG. 44A****FIG. 44B****FIG. 44C**

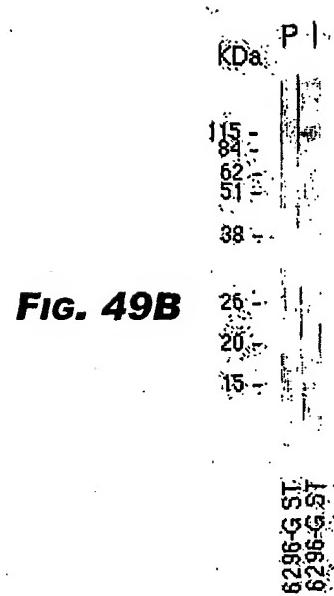
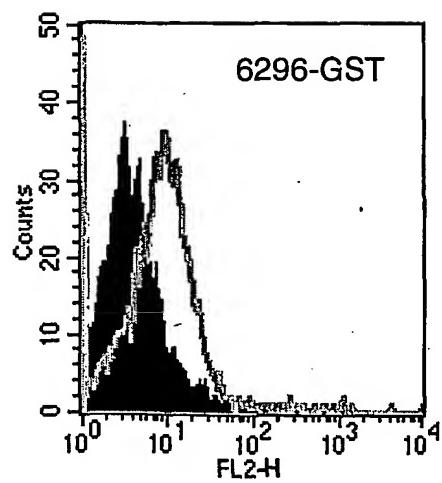
47/169

FIGURE 47**FIG. 47A****FIG. 47B****FIG. 47C**

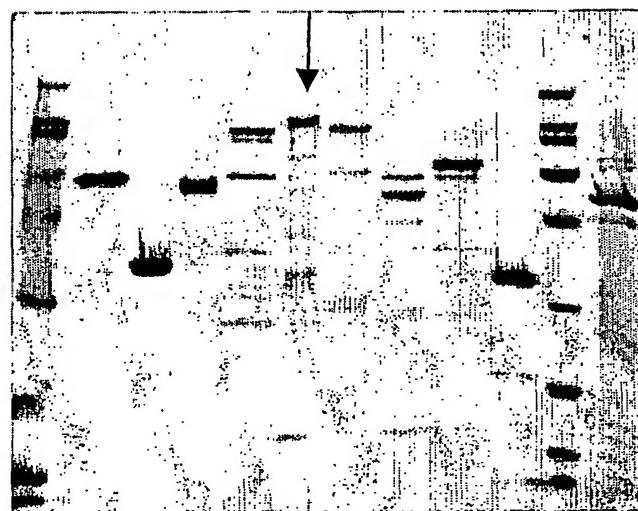
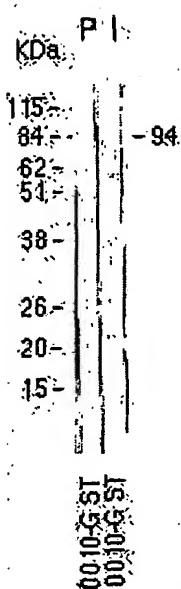
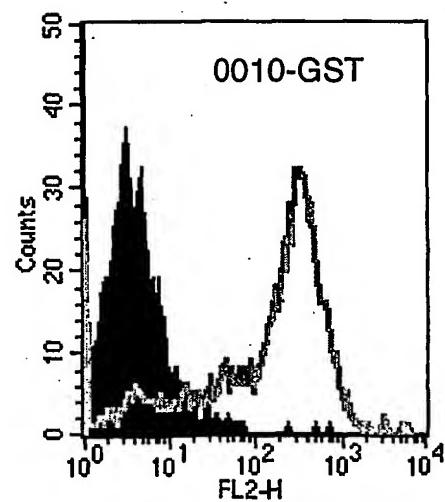
46/169

FIGURE 46**FIG. 46A****FIG. 46B**

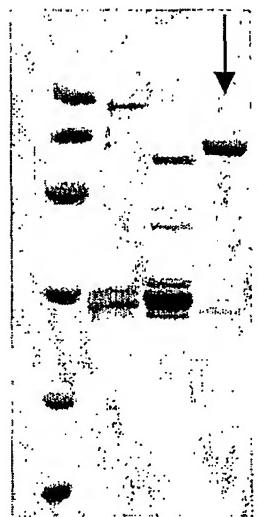
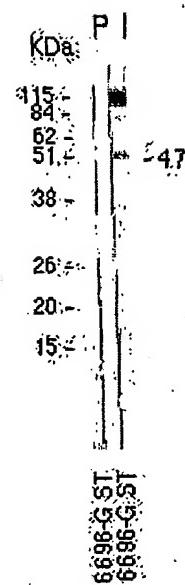
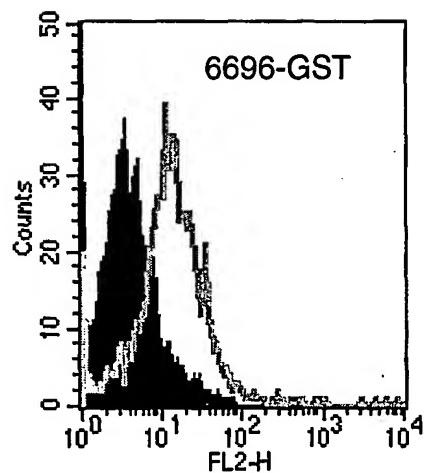
49/169

FIGURE 49**FIG. 49A****FIG. 49B****FIG. 49C**

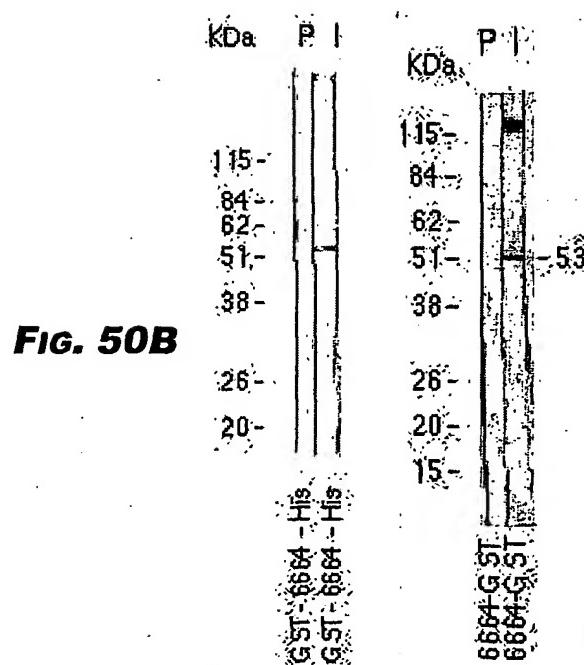
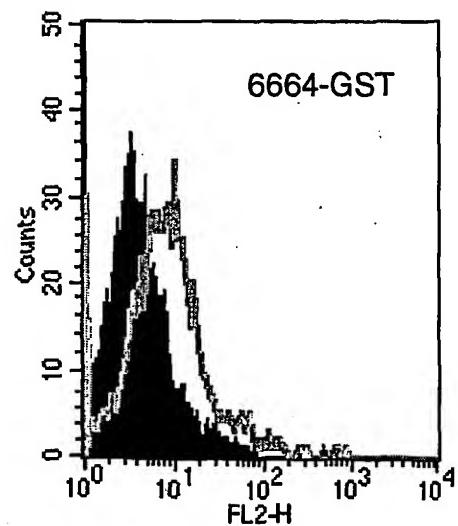
48/169

FIGURE 48**FIG. 48A****FIG. 48B****FIG. 48C**

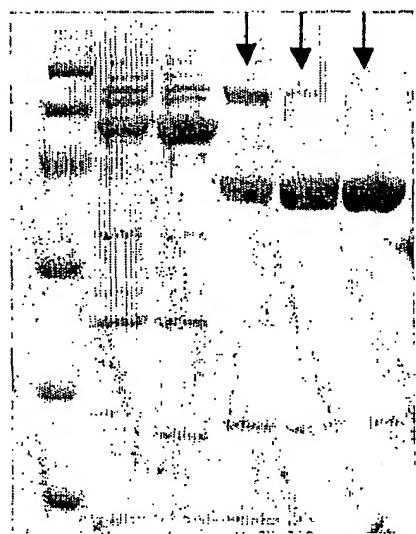
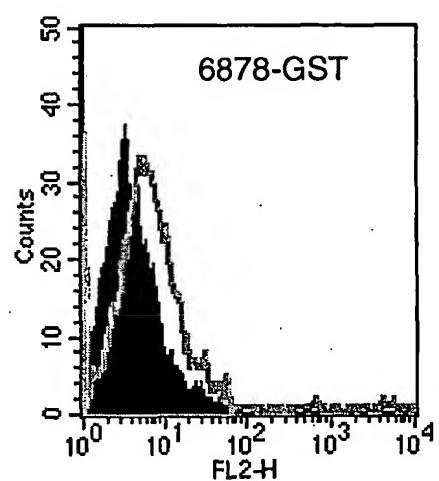
51/169

FIGURE 51**FIG. 51A****FIG. 51B****FIG. 51C**

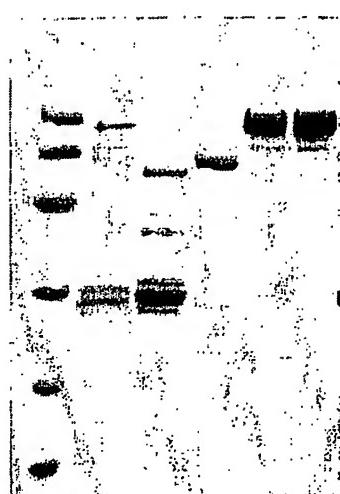
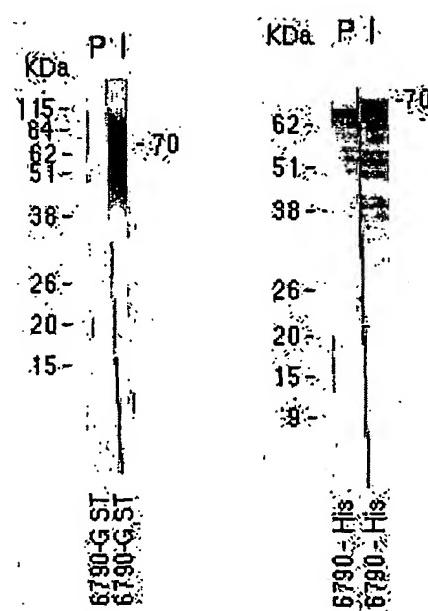
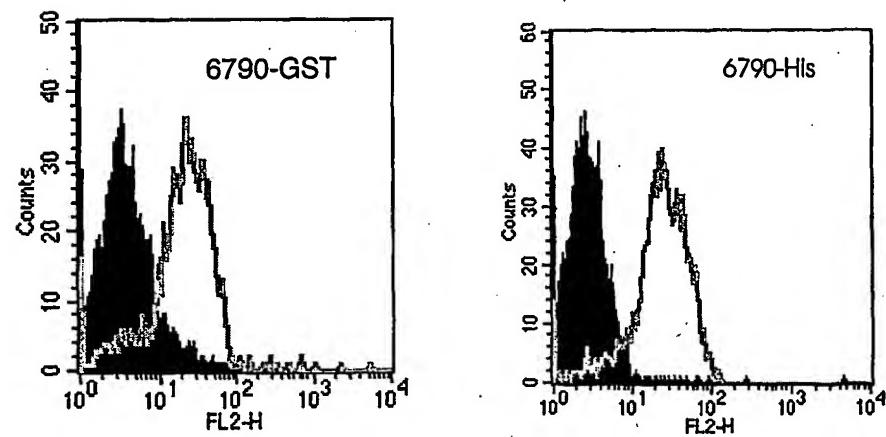
50/169

FIGURE 50**FIG. 50A****FIG. 50B****FIG. 50C**

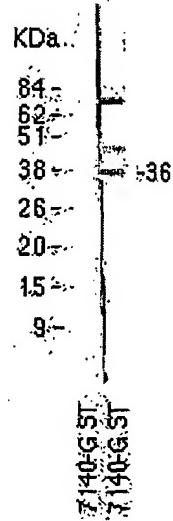
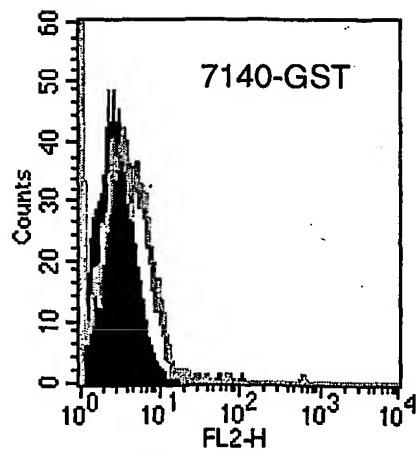
53/169

FIGURE 53**FIG. 53A****FIG. 53B**

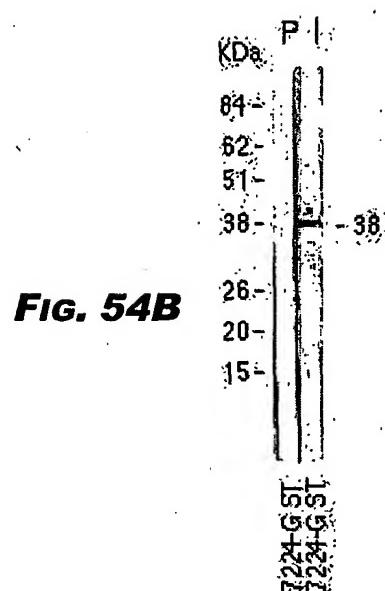
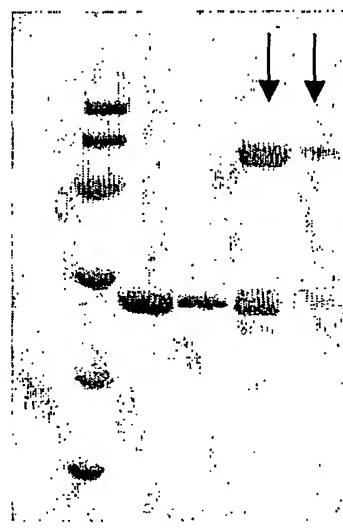
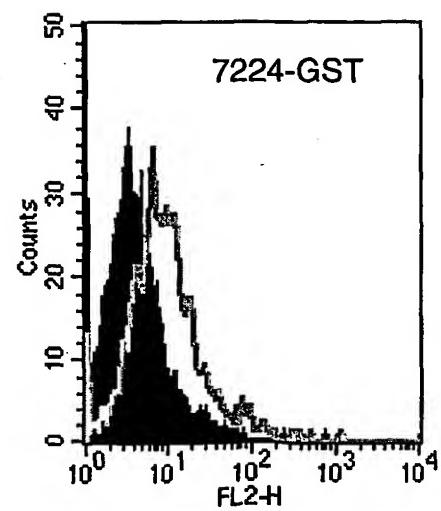
52/169

FIGURE 52**FIG. 52A****FIG. 52B****FIG. 52C**

55/169

FIGURE 55**FIG. 55A****FIG. 55B****FIG. 55C**

54/169

FIGURE 54**FIG. 54A****FIG. 54B****FIG. 54C**

57/169

FIGURE 57

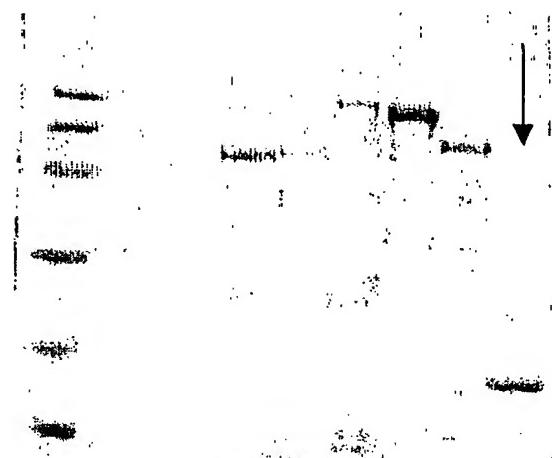


FIG. 57A

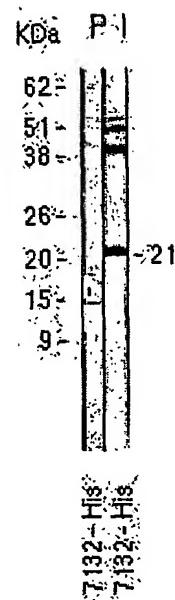


FIG. 57B

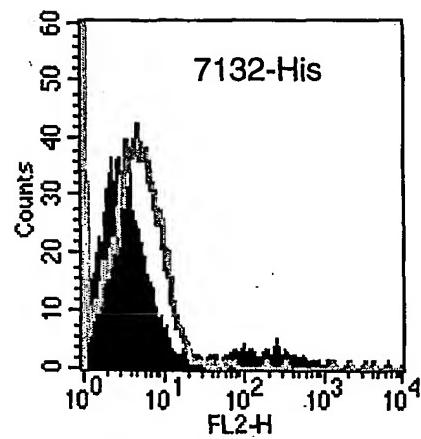
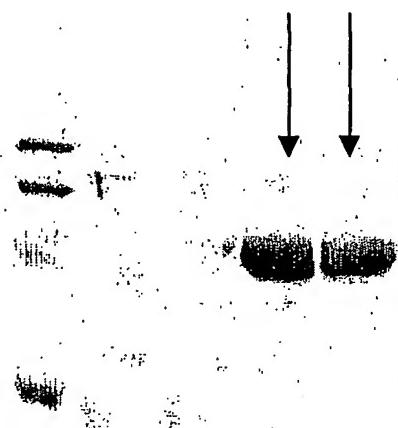
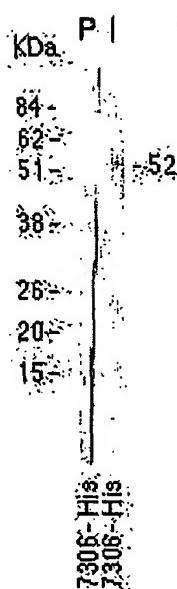
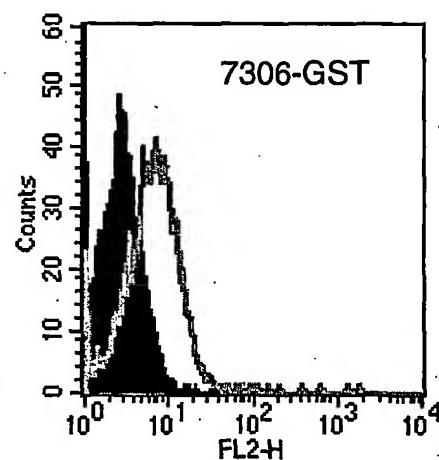
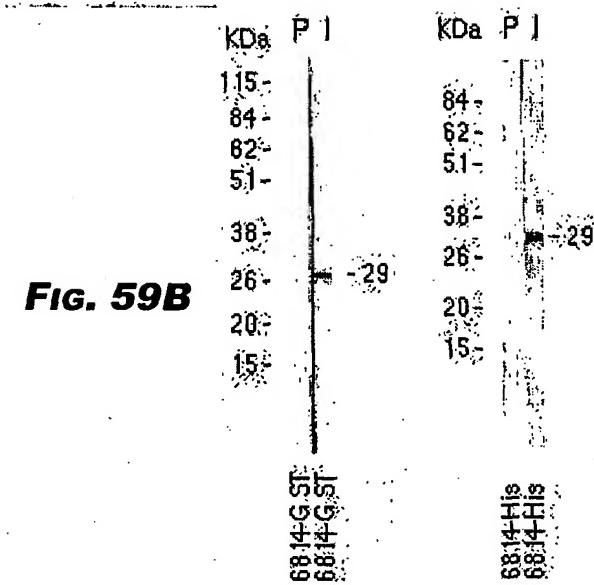
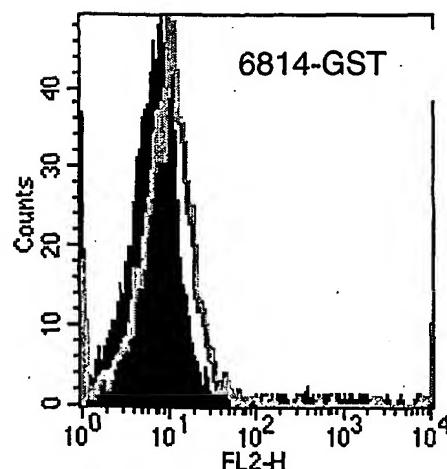


FIG. 57C

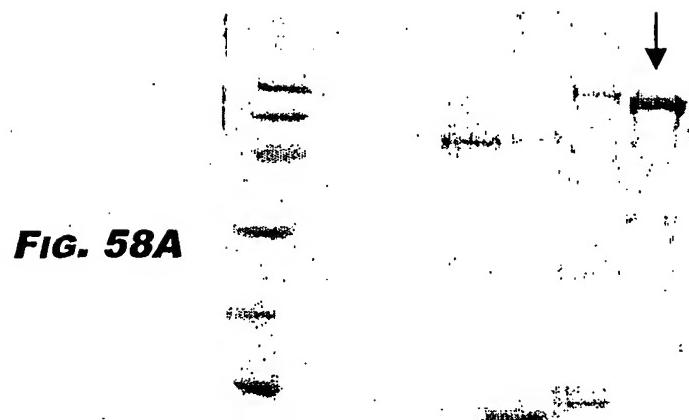
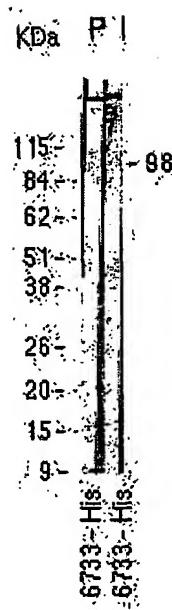
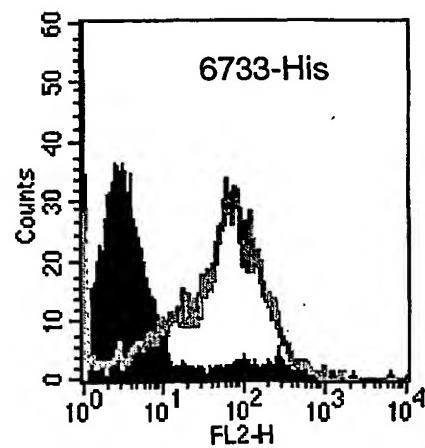
56/169

FIGURE 56**FIG. 56A****FIG. 56B****FIG.
56C****FIG.
56D**

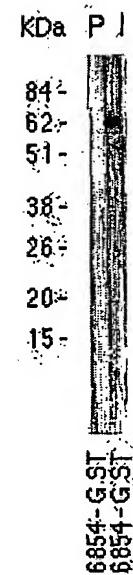
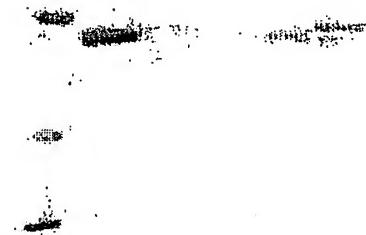
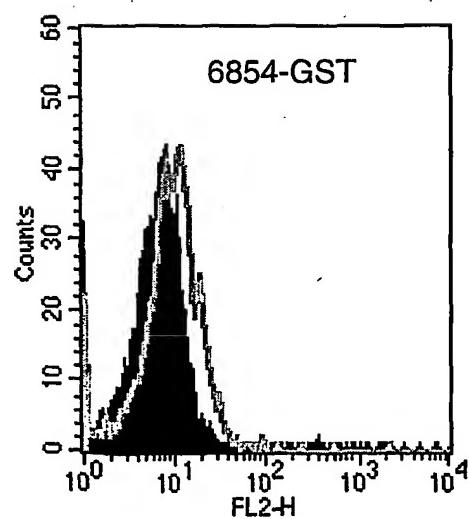
59/169

FIGURE 59**FIG. 59A****FIG. 59B****FIG. 59C**

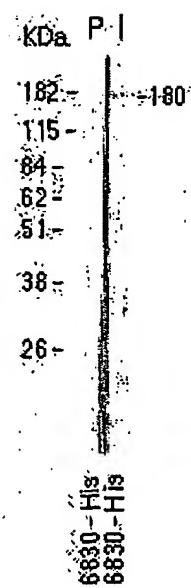
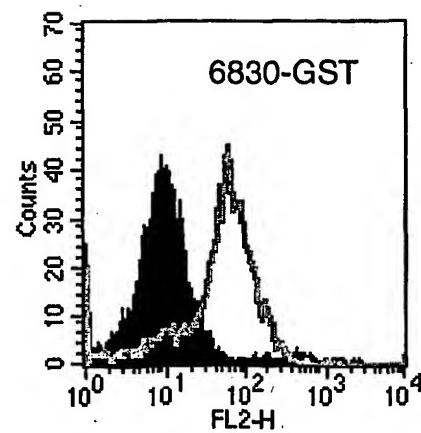
58/169

FIGURE 58**FIG. 58A****FIG. 58B****FIG. 58C**

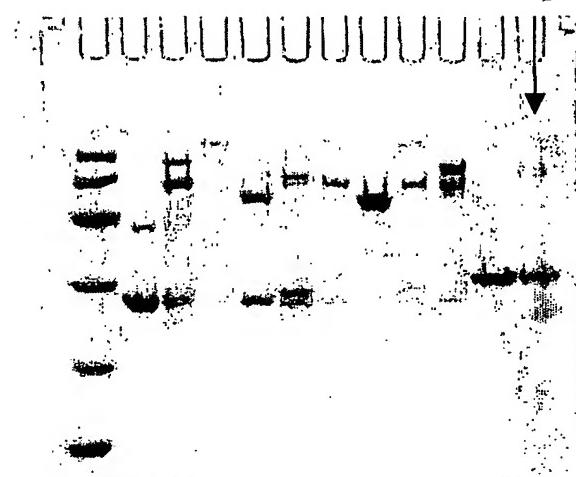
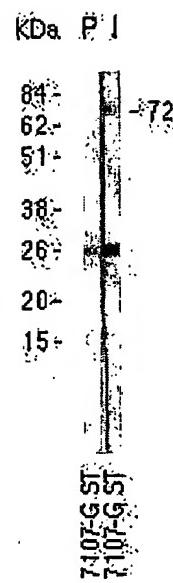
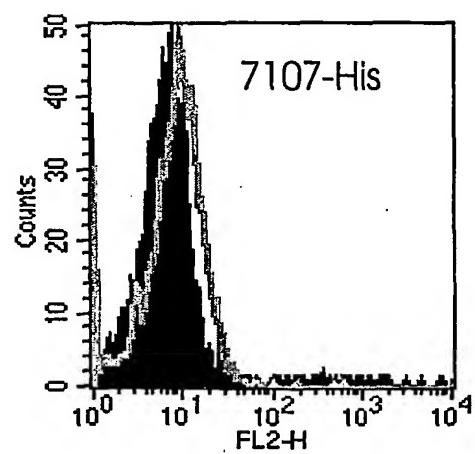
61/169

FIGURE 61**FIG. 61A****FIG. 61B****FIG. 61C**

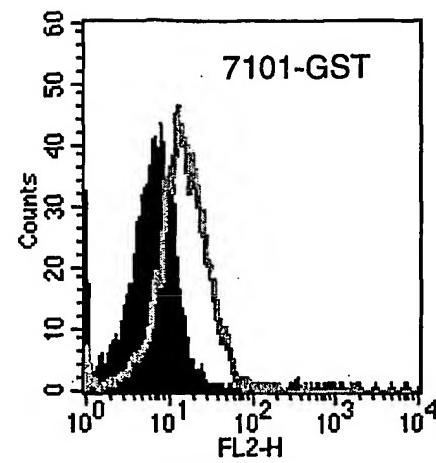
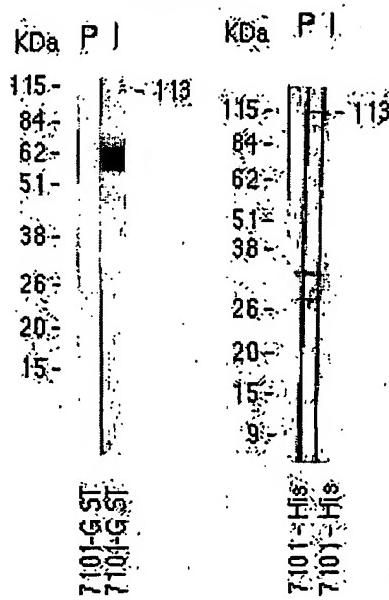
60/169

FIGURE 60**FIG. 60A****FIG. 60B****FIG. 60C**

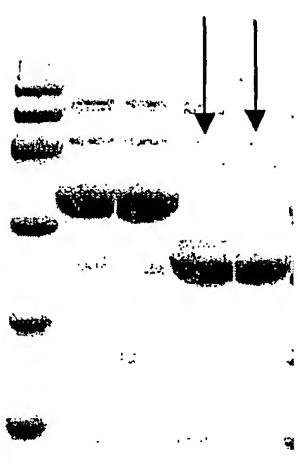
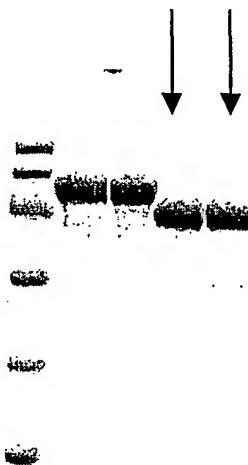
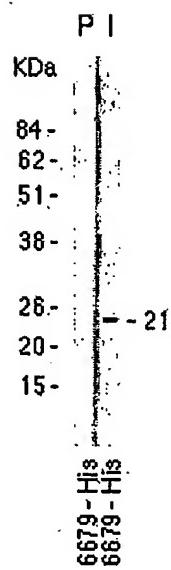
63/169

FIGURE 63**FIG. 63A****FIG. 63B****FIG. 63C**

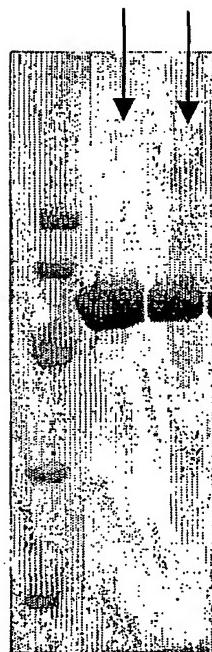
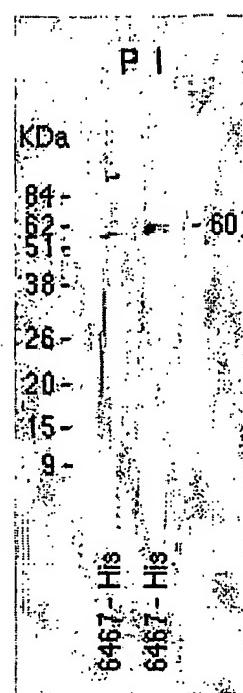
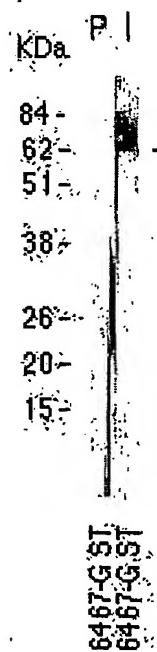
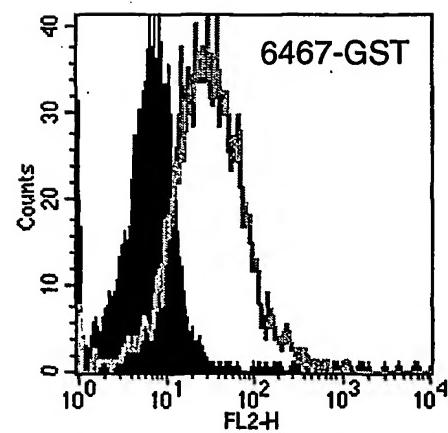
62/169

FIGURE 62**FIG. 62A****FIG. 62C****FIG. 62B**

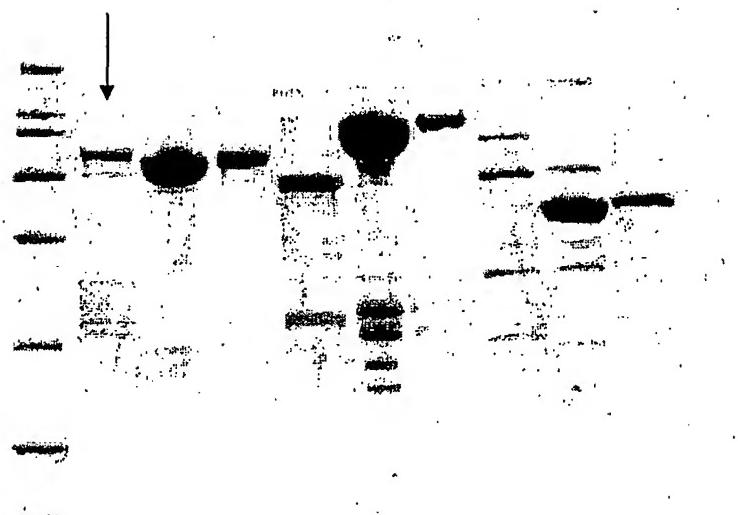
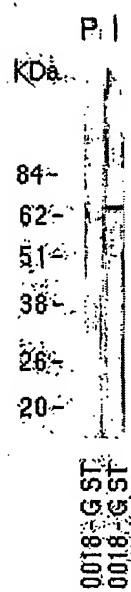
65/169

FIGURE 65**FIG. 65A****FIG. 65B****FIG. 65C**

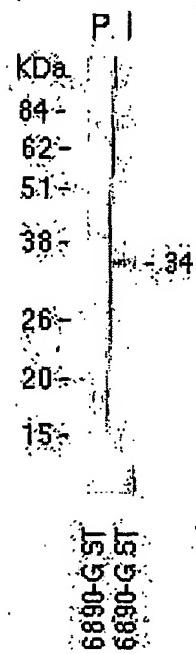
64/169

FIGURE 64**FIG. 64A****FIG. 64B****FIG. 64C****FIG. 64D**

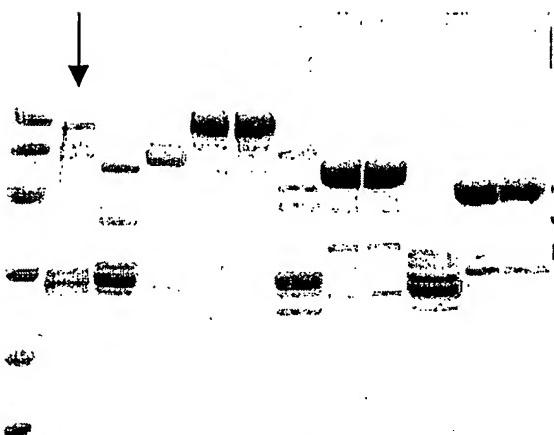
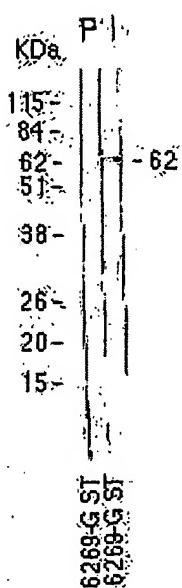
67/169

FIGURE 67**FIG. 67A****FIG. 67B**

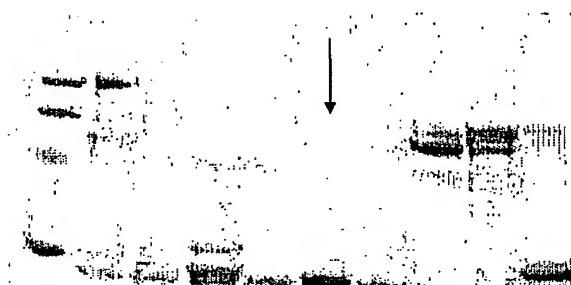
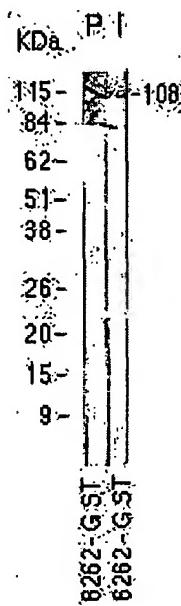
66/169

FIGURE 66**FIG. 66A****FIG. 66B**

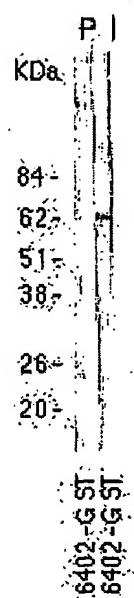
69/169

FIGURE 69**FIG. 69A****FIG. 69B**

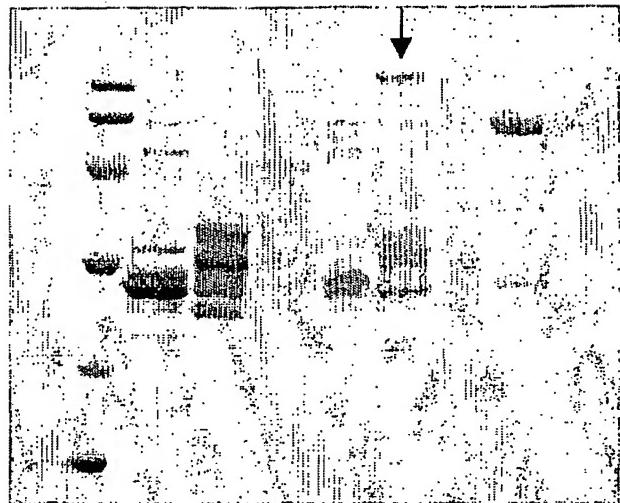
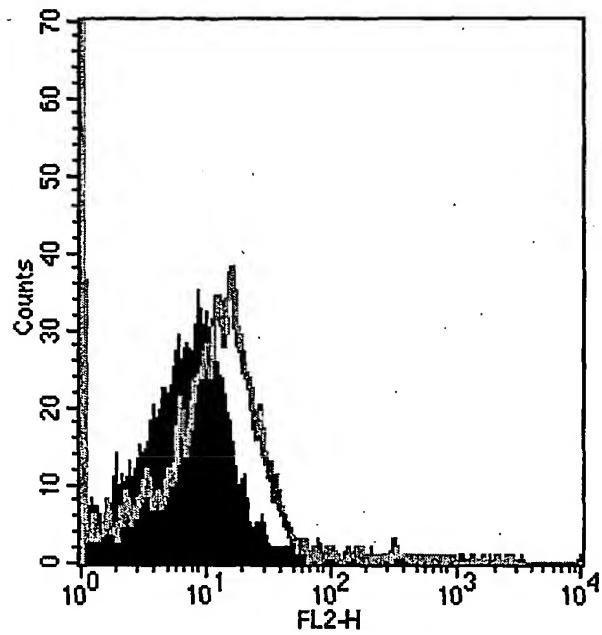
68/169

FIGURE 68**FIG. 68A****FIG. 68B**

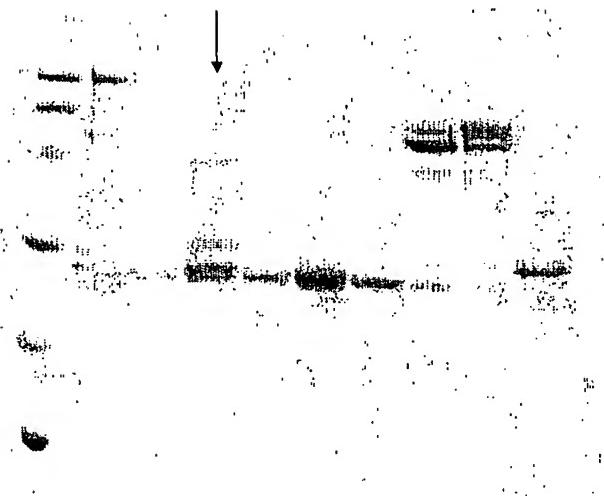
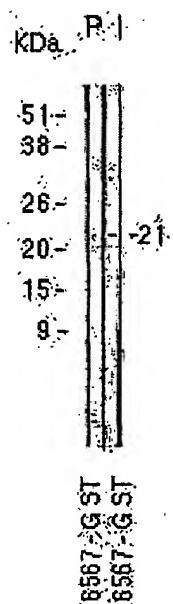
71/169

FIGURE 71**FIG. 71A****FIG. 71B**

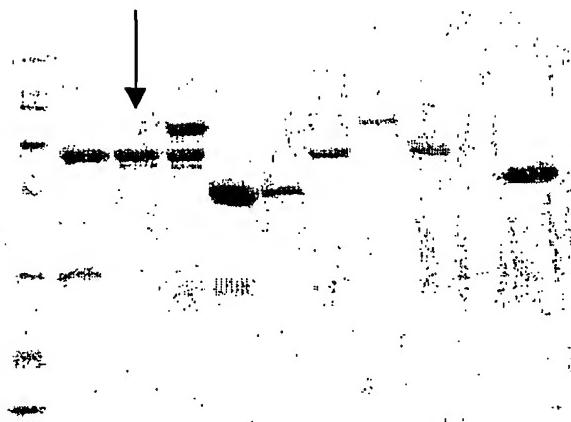
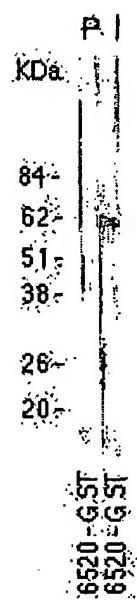
70/169

FIGURE 70**FIG. 70A****FIG. 70B**

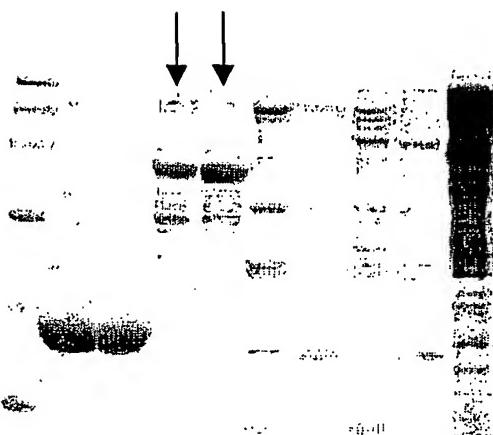
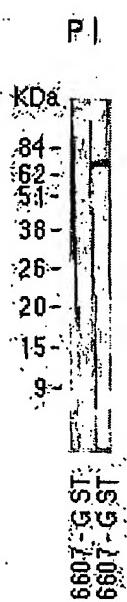
73/169

FIGURE 73**FIG. 73A****FIG. 73B**

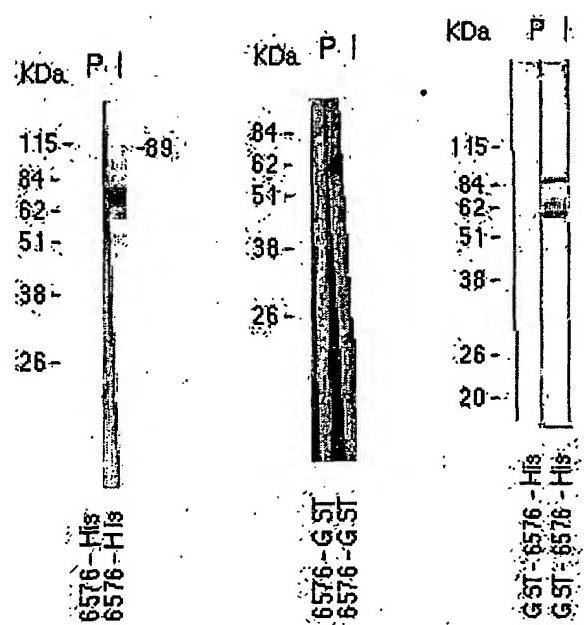
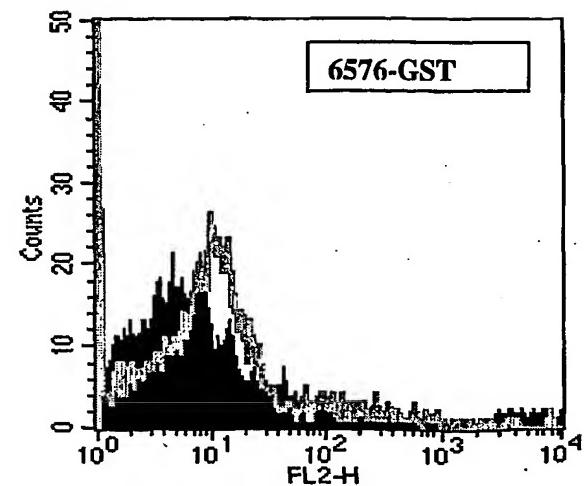
72/169

FIGURE 72**FIG. 72A****FIG. 72B**

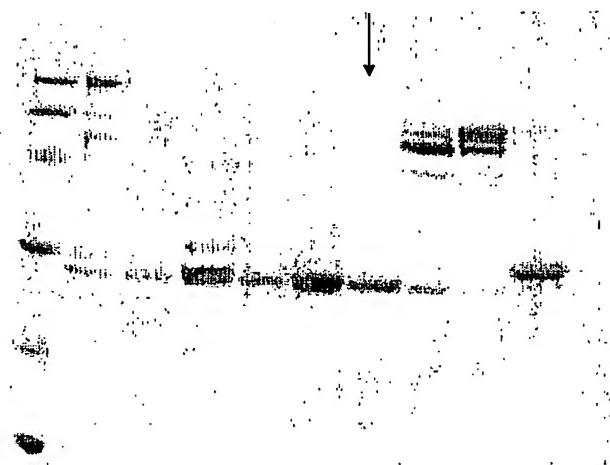
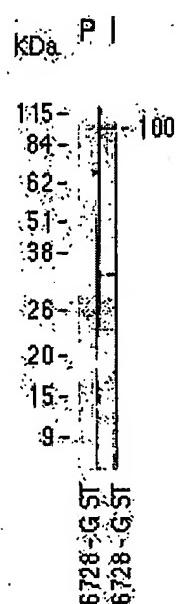
75/169

FIGURE 75**FIG. 75A****FIG. 75B**

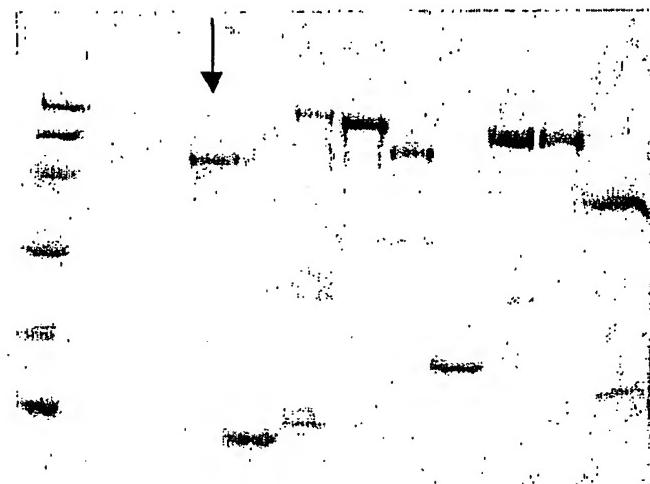
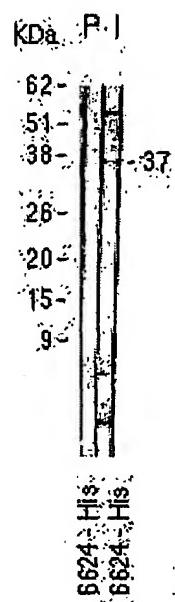
74/169

FIGURE 74**FIG. 74A****FIG. 74B****FIG. 74C**

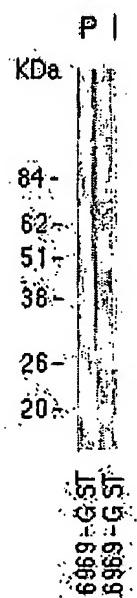
77/169

FIGURE 77**FIG. 77A****FIG. 77B**

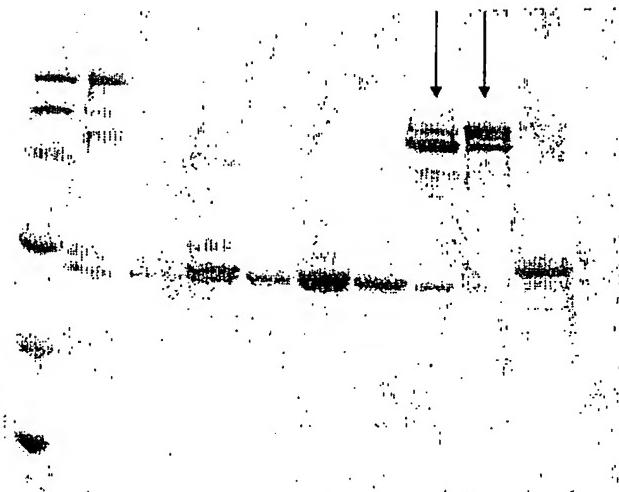
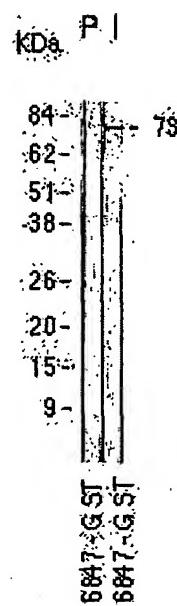
76/169

FIGURE 76**FIG. 76A****FIG. 76B**

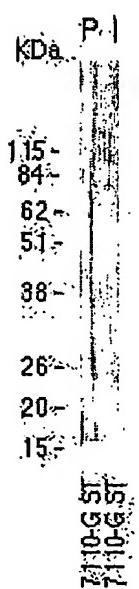
79/169

FIGURE 79**FIG. 79A****FIG. 79B**

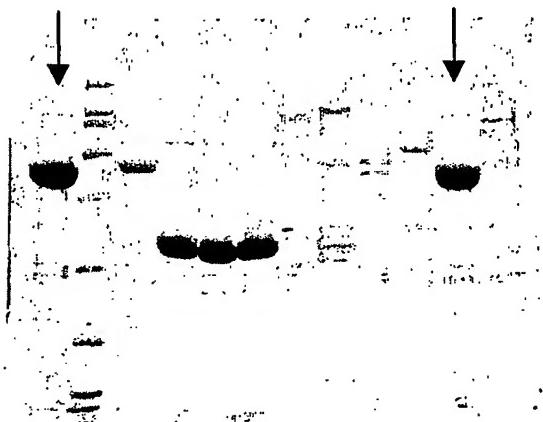
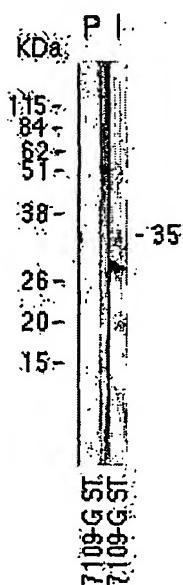
78/169

FIGURE 78**FIG. 78A****FIG. 78B**

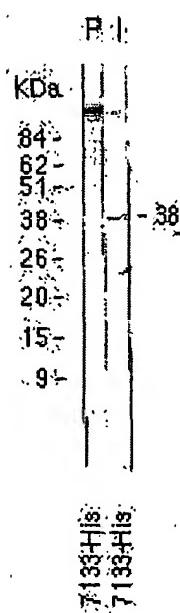
81/169

FIGURE 81**FIG. 81A****FIG. 81B**

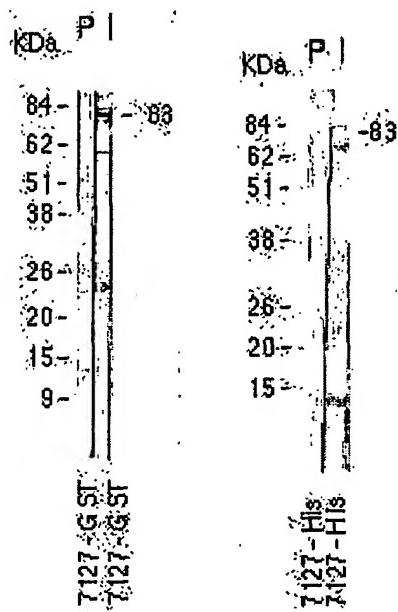
80/169

FIGURE 80**FIG. 80A****FIG. 80B**

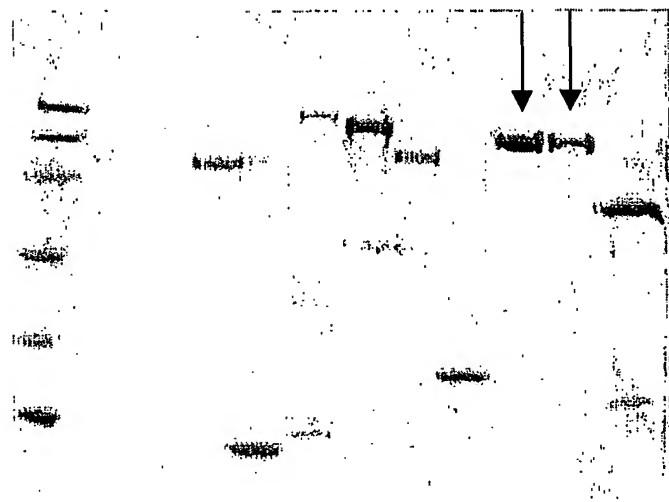
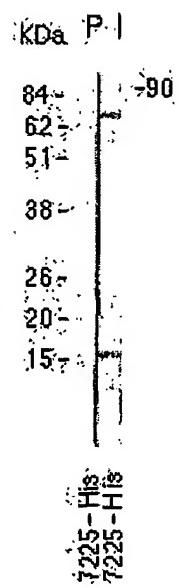
83/169

FIGURE 83**FIG. 83A****FIG. 83B**

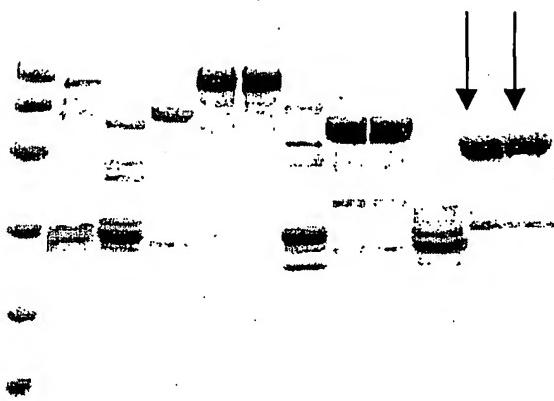
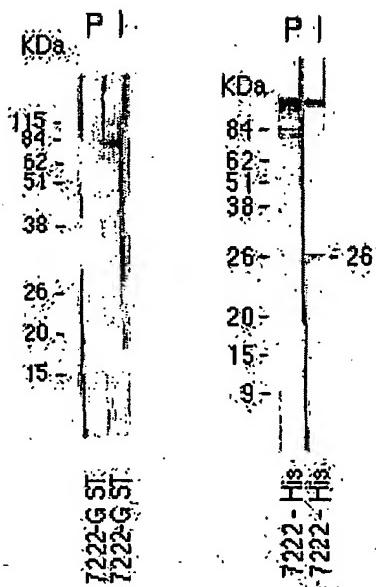
82/169

FIGURE 82**FIG. 82A****FIG. 82B**

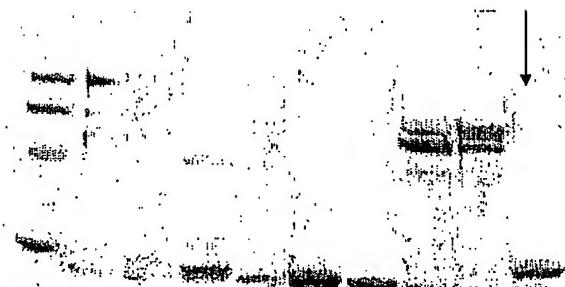
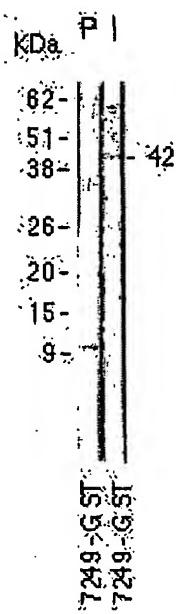
85/169

FIGURE 85**FIG. 85A****FIG. 85B**

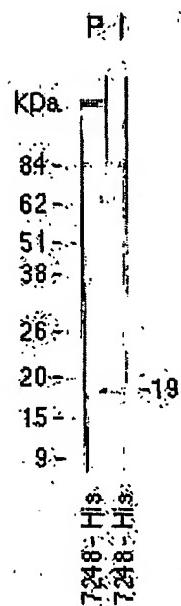
84/169

FIGURE 84**FIG. 84A****FIG. 84B**

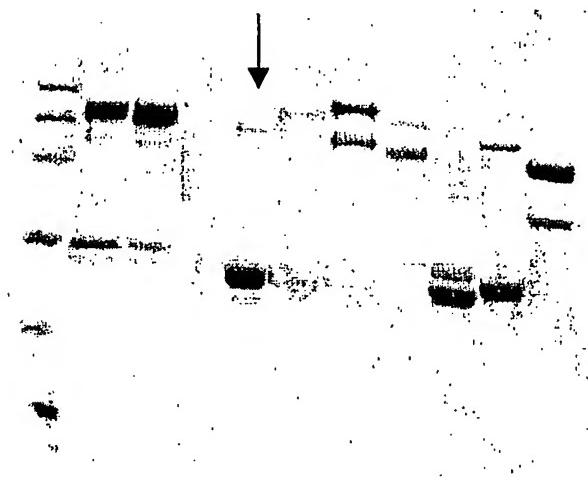
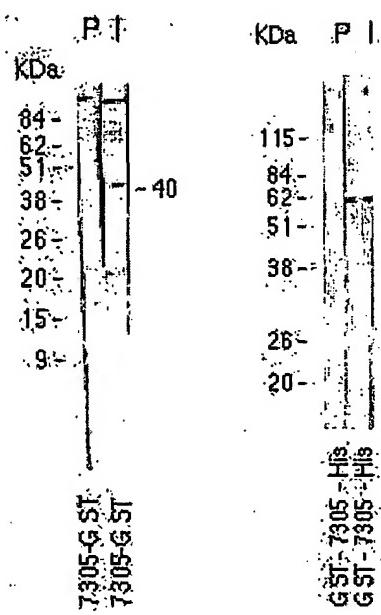
87/169

FIGURE 87**FIG. 87A****FIG. 87B**

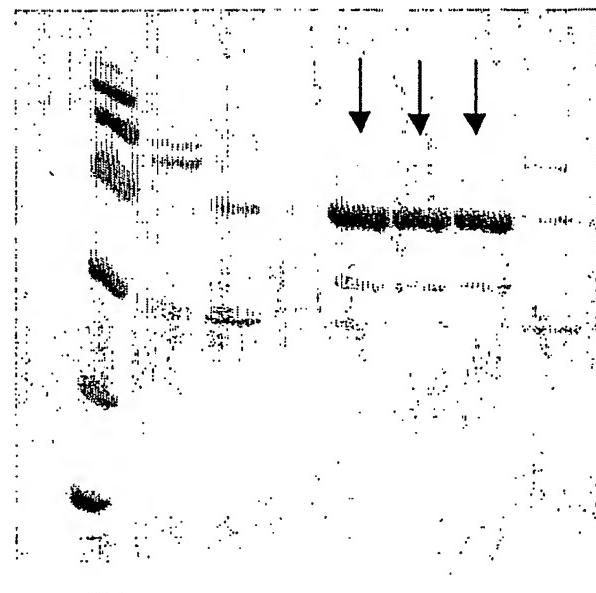
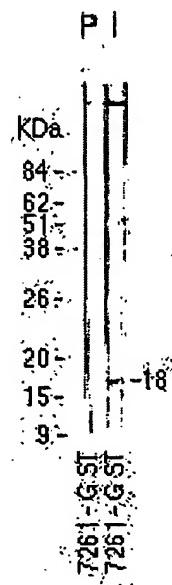
86/169

FIGURE 86**FIG. 86A****FIG. 86B**

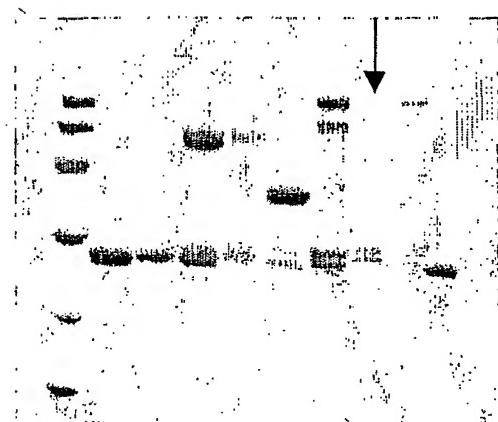
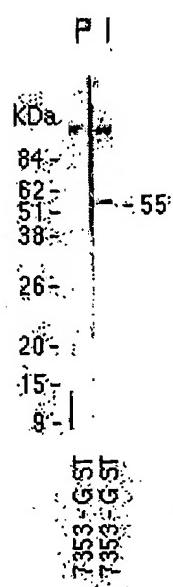
89/169

FIGURE 89**FIG. 89A****FIG. 89B**

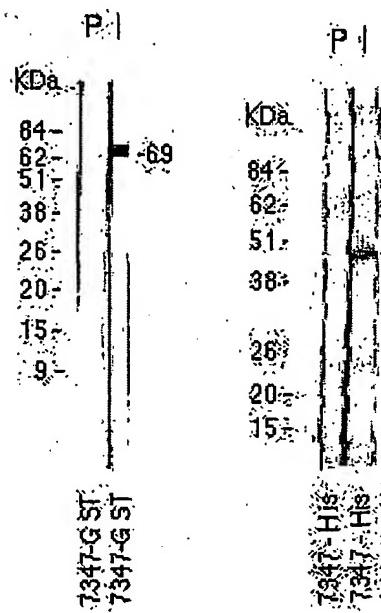
88/169

FIGURE 88**FIG. 88A****FIG. 88B**

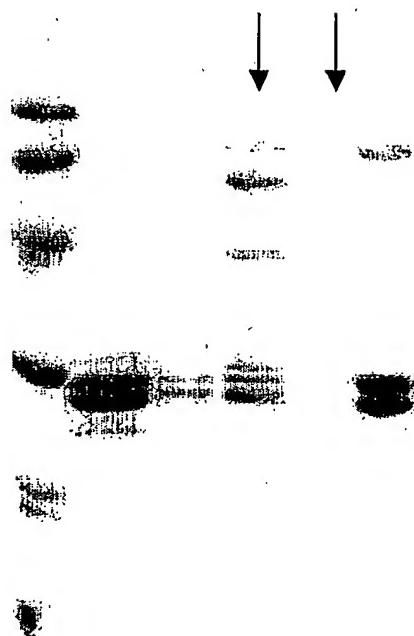
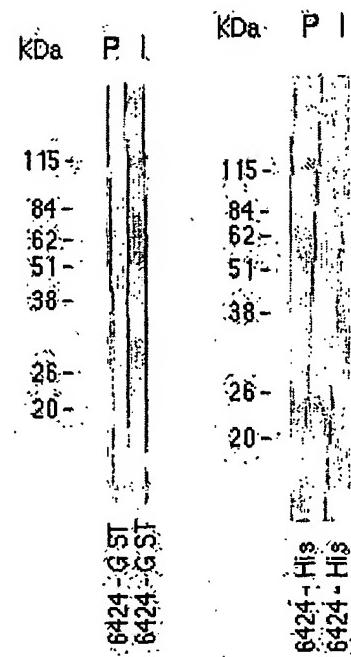
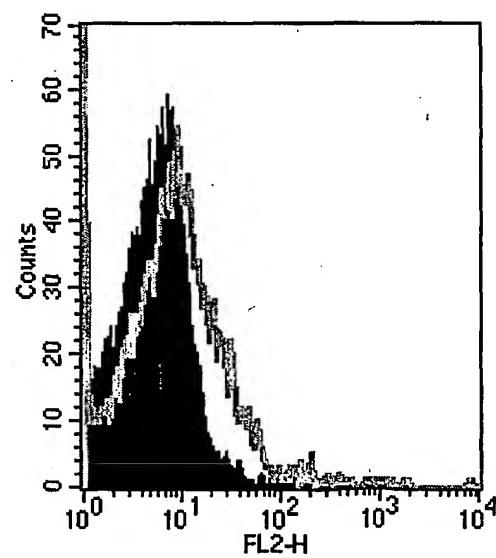
91/169

FIGURE 91**FIG. 91A****FIG. 91B**

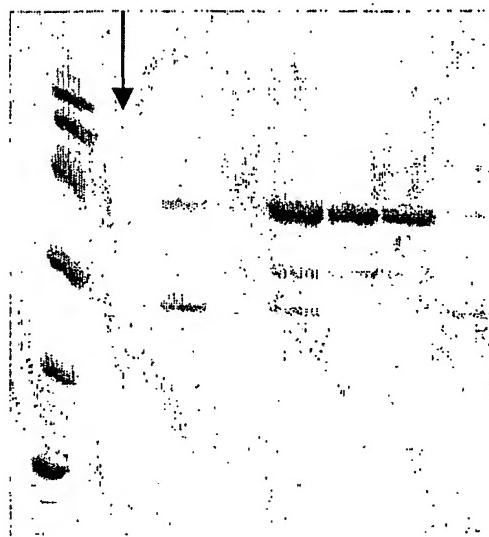
90/169

FIGURE 90**FIG. 90A****FIG. 90B**

93/169

FIGURE 93***FIG. 93A******FIG. 93B******FIG. 93C***

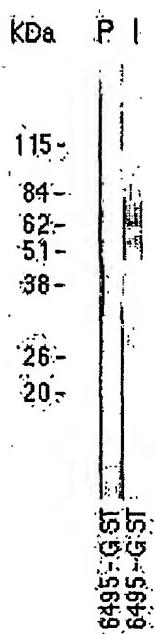
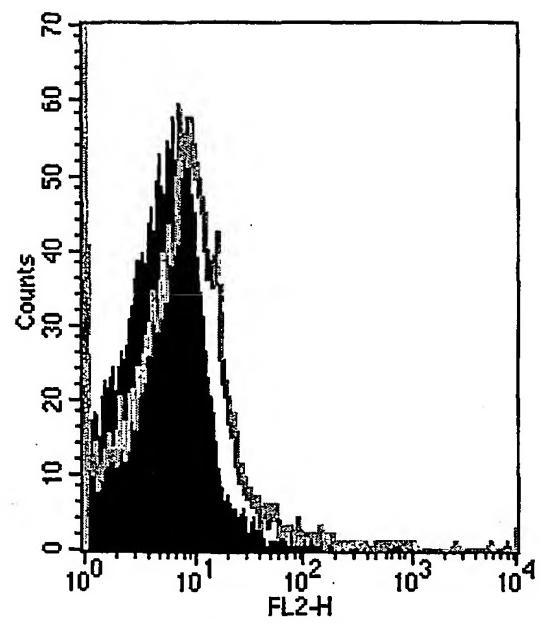
92/169

FIGURE 92**FIG. 92A**

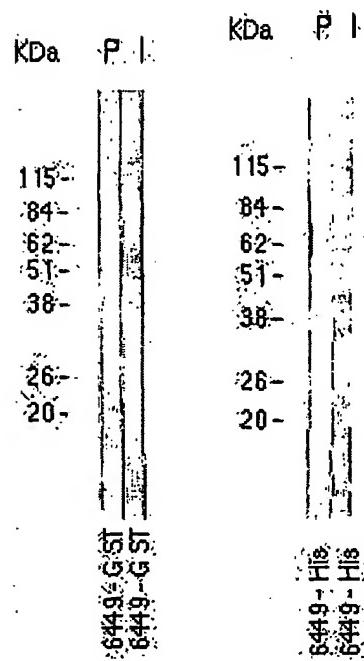
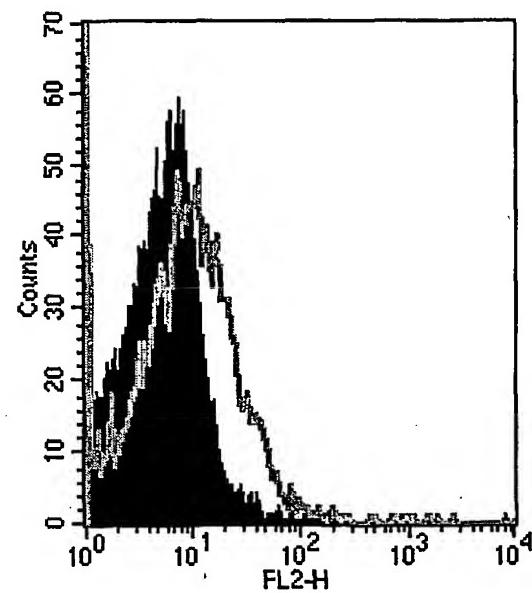
KDa P.I.

84
62
51
36
26
20
15**FIG. 92B**7408-His
7408-His

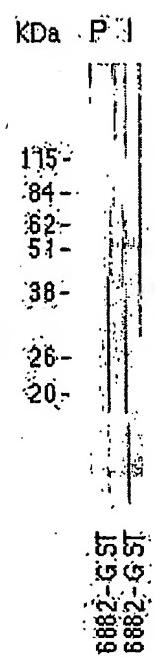
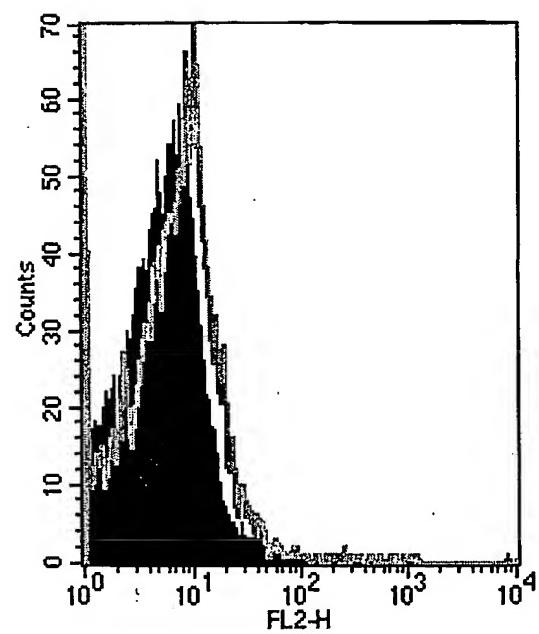
95/169

FIGURE 95**FIG. 95A****FIG. 95B****FIG. 95C**

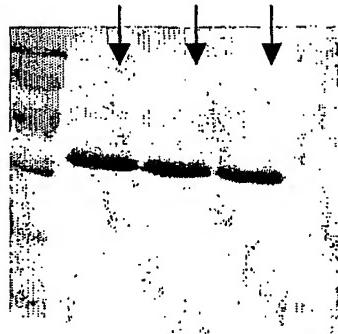
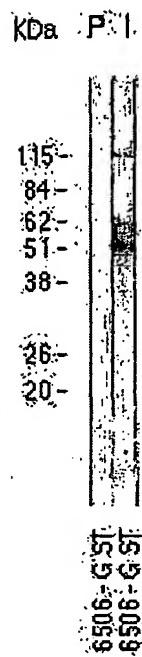
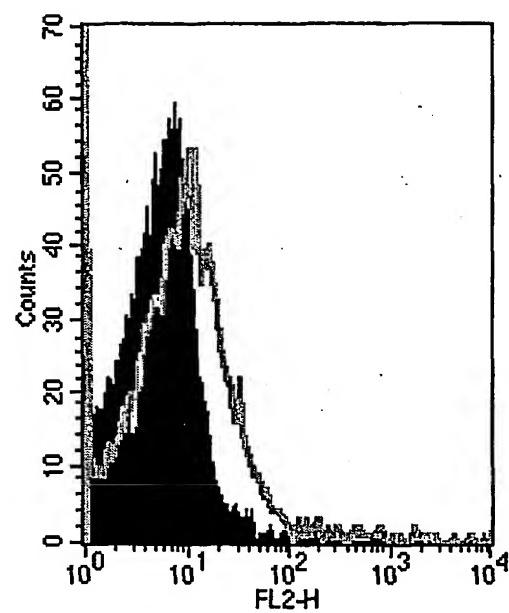
94/169

FIGURE 94**FIG. 94A****FIG. 94B****FIG. 94C**

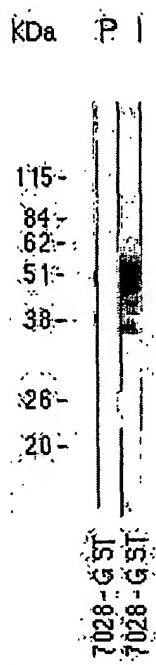
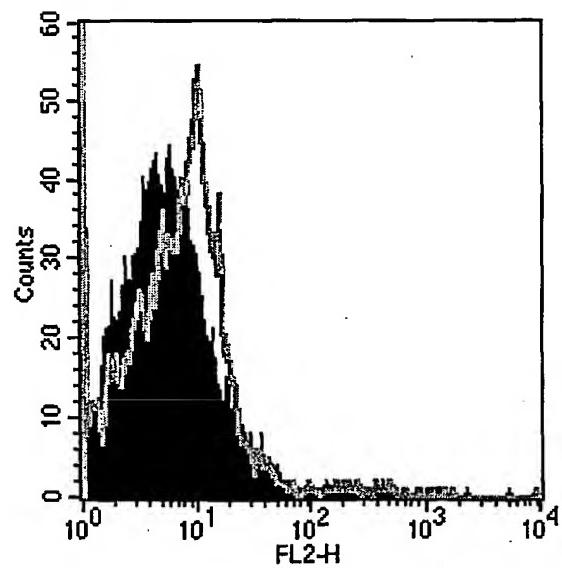
97/169

FIGURE 97**FIG. 97A****FIG. 97B****FIG. 97C**

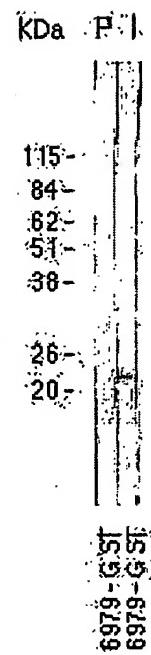
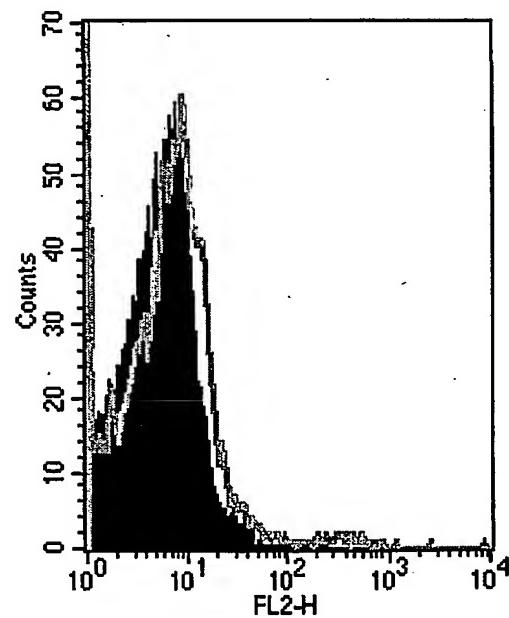
96/169

FIGURE 96**FIG.
96A****FIG.
96B****FIG.
96C****FIG. 96D**

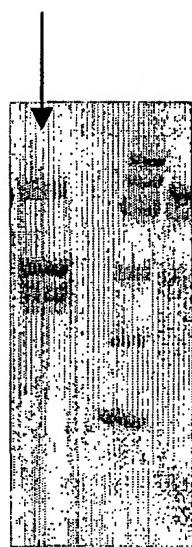
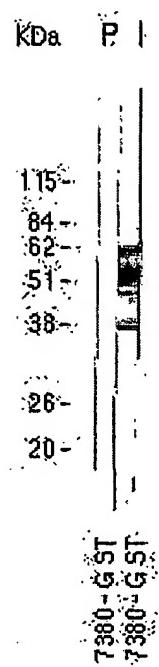
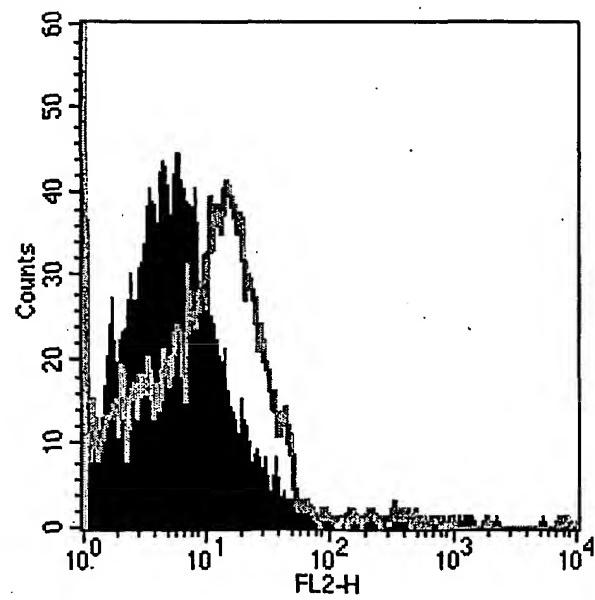
99/169

FIGURE 99**FIG. 99A****FIG. 99B****FIG. 99C**

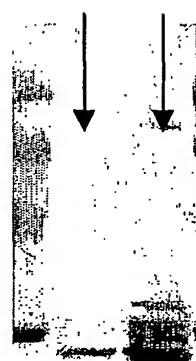
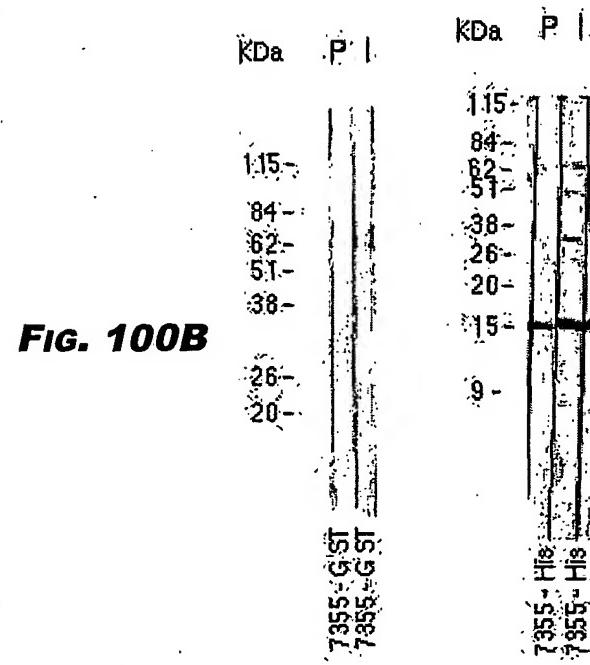
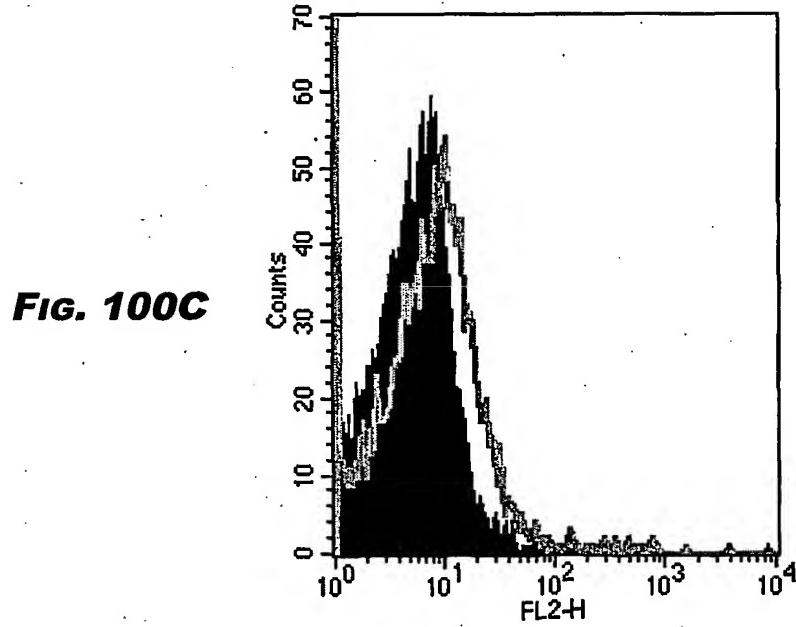
98/169

FIGURE 98**FIG. 98A****FIG. 98B****FIG. 98C**

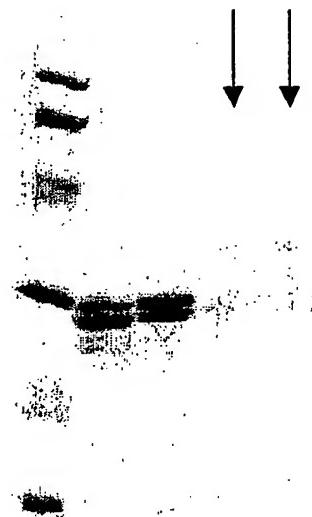
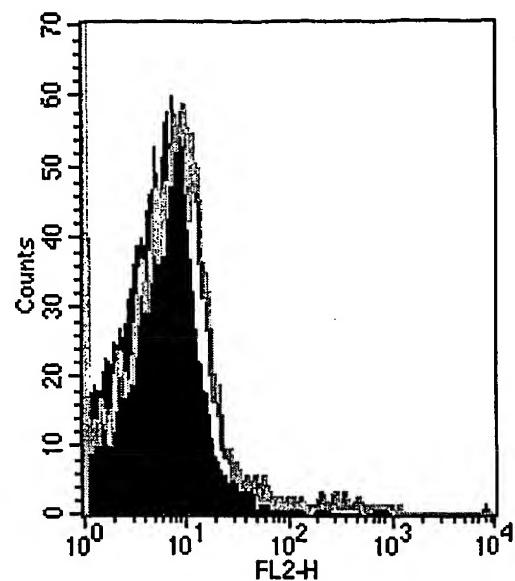
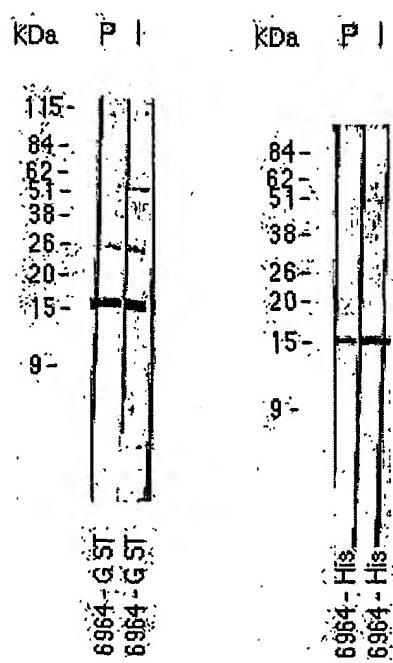
101/169

FIGURE 101**FIG. 101A****FIG. 101B****FIG. 101C**

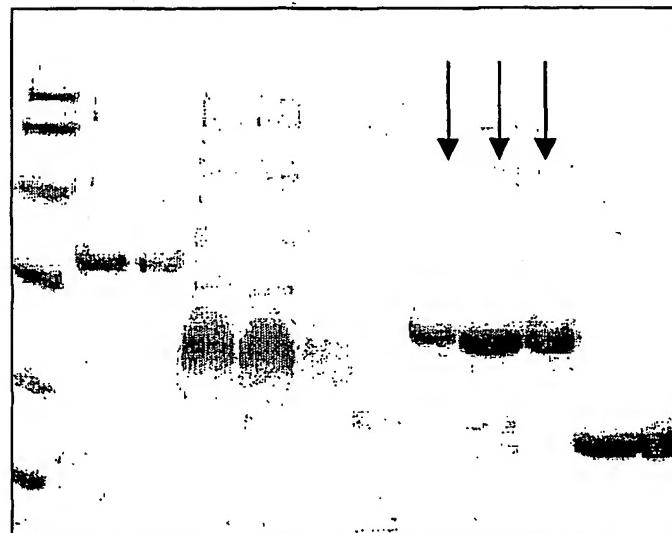
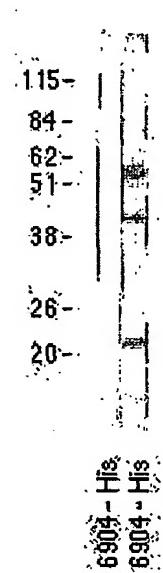
100/169

FIGURE 100**FIG. 100A****FIG. 100B****FIG. 100C**

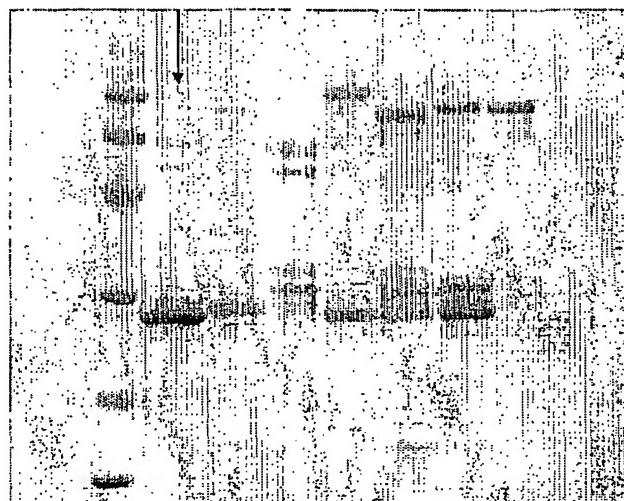
103/169

FIGURE 103**FIG.
103A****FIG.
103C****FIG. 103B**

102/169

FIGURE 102**FIG. 102A****FIG. 102B**

105/169

FIGURE 105**FIG. 105A**

KDa P L

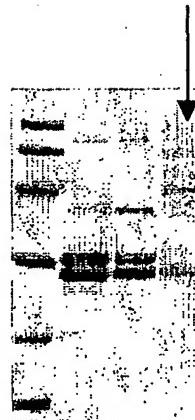
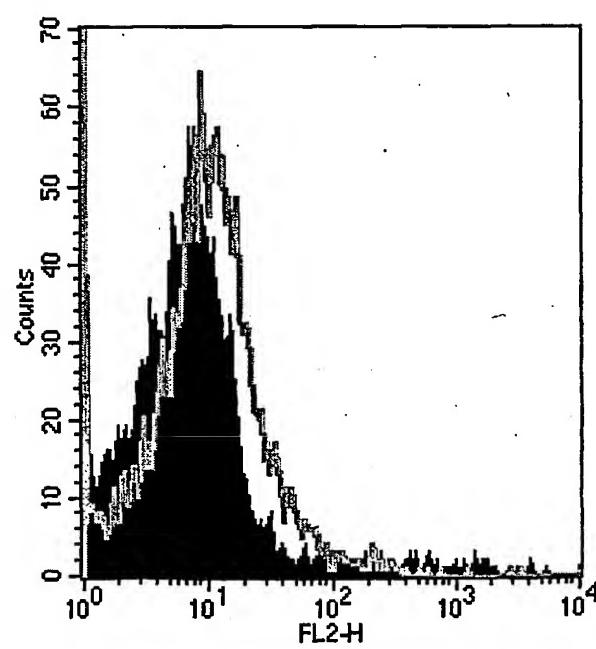
115-
84-
62-
51-
38-

FIG. 105B

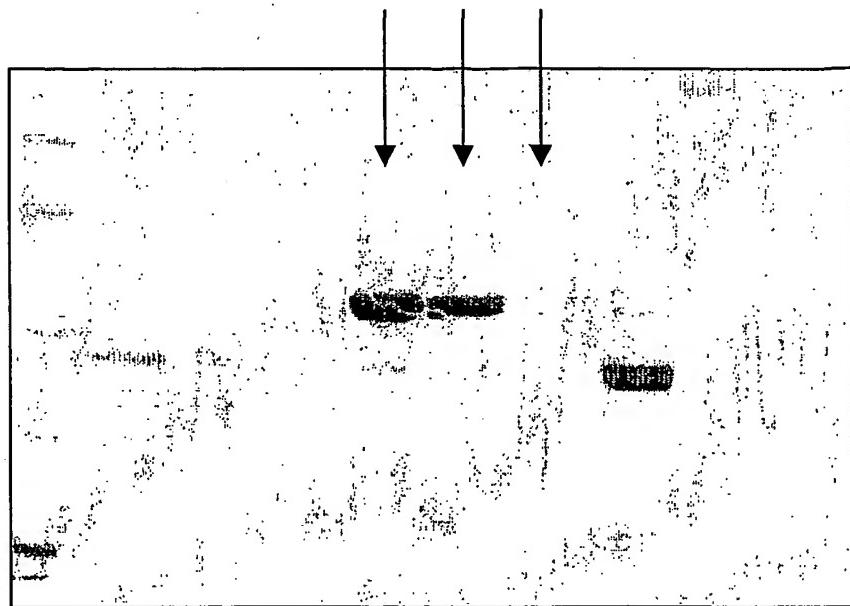
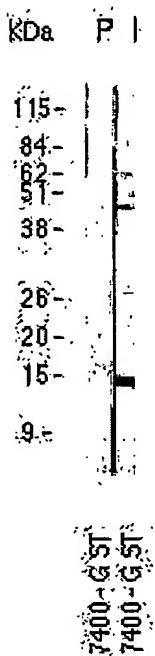
26-
20-

6281-G ST
6281-G ST

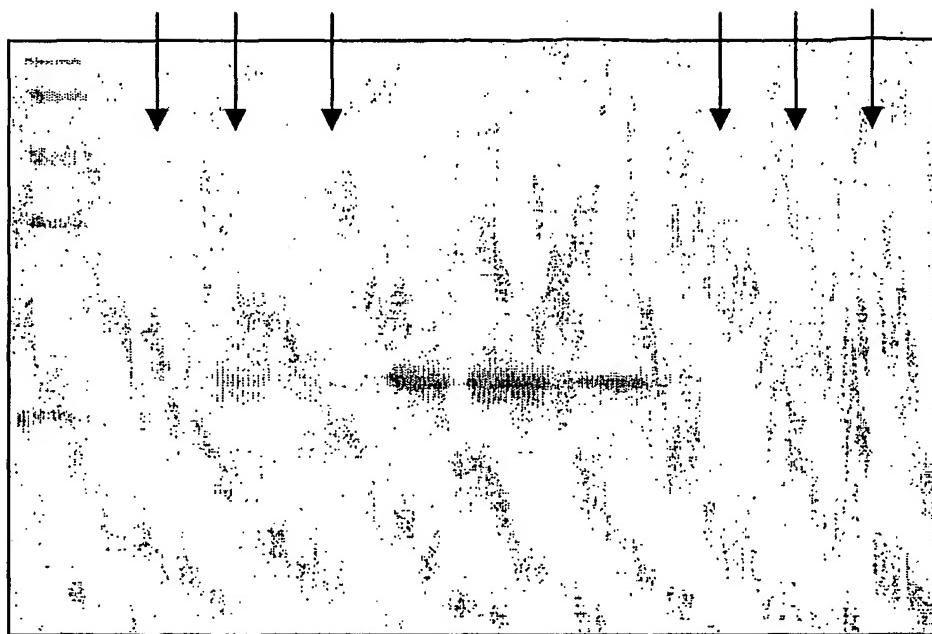
104/169

FIGURE 104**FIG. 104A****FIG. 104B****FIG. 104C**

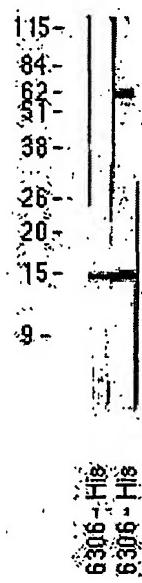
107/169

FIGURE 108**FIG. 108A****FIG. 108B**

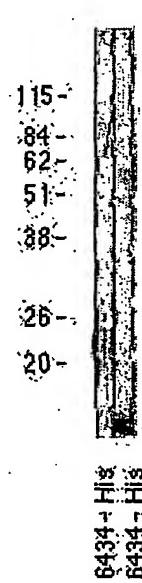
106/169

FIGURE 106**FIG. 106A****FIG. 106B**

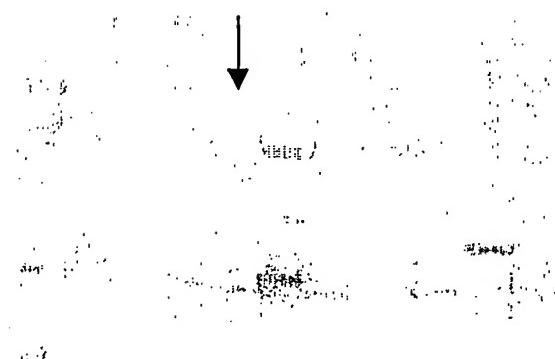
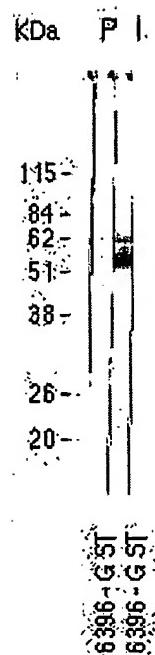
KDa P I.

**FIGURE 107**

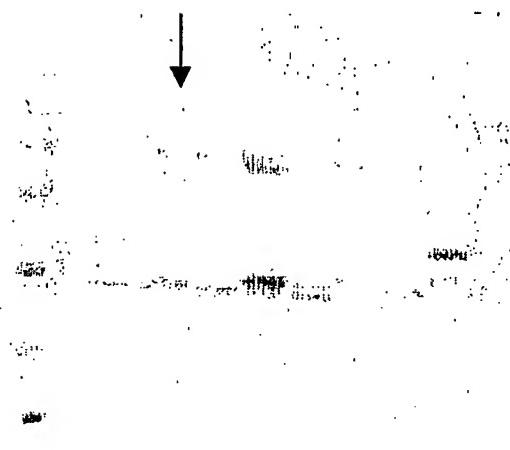
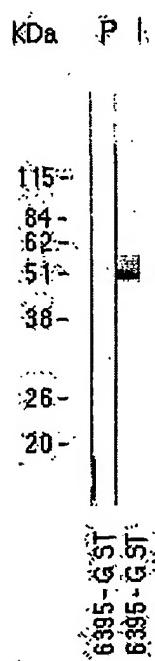
KDa P I.



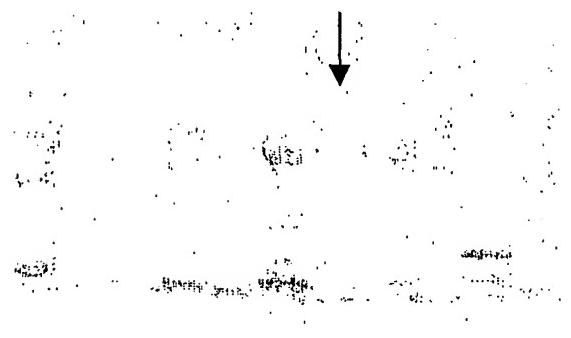
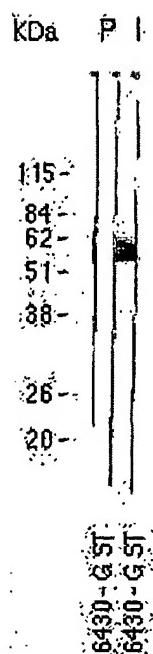
109/169

FIGURE 110**FIG. 110A****FIG. 110B**

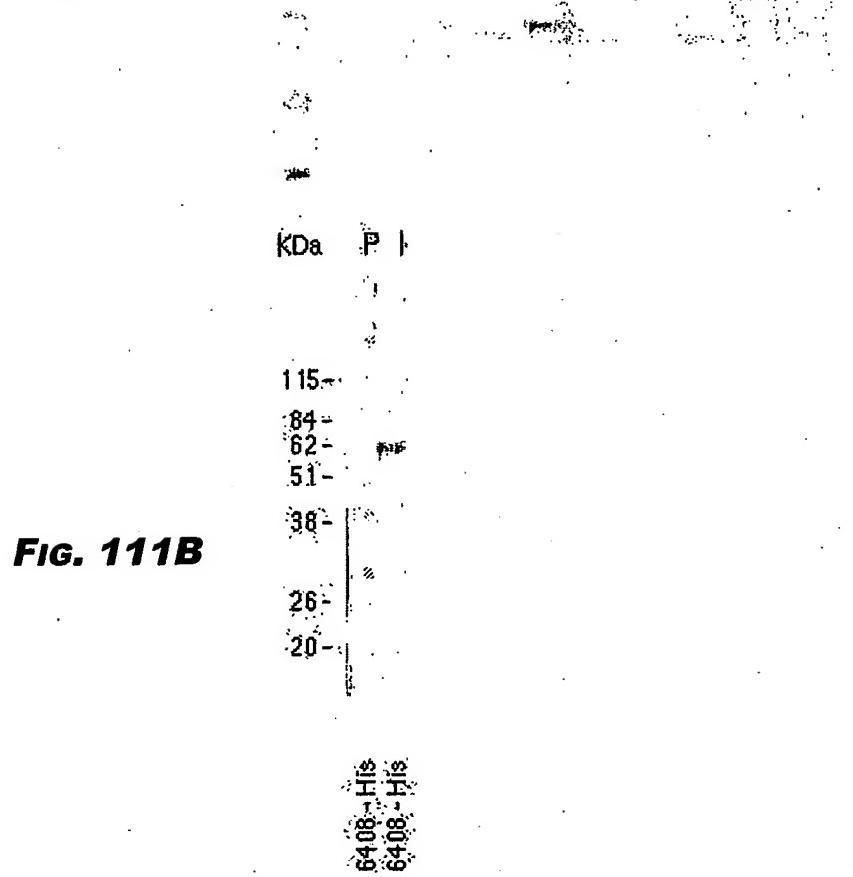
108/169

FIGURE 109**FIG. 109A****FIG. 109B**

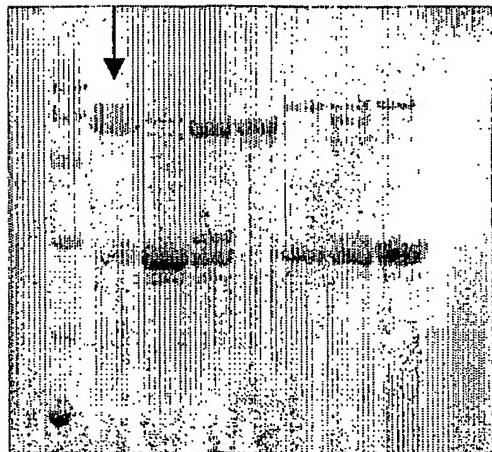
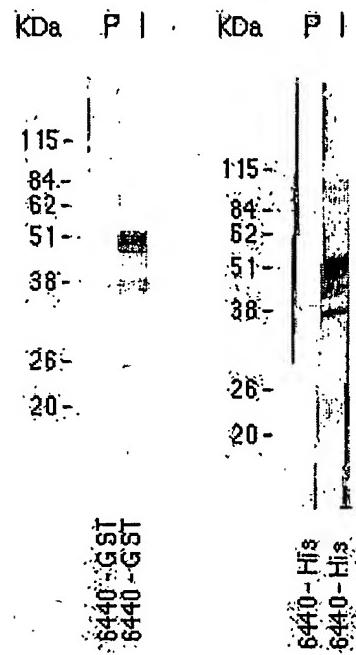
111/169

FIGURE 112**FIG. 112A****FIG. 112B**

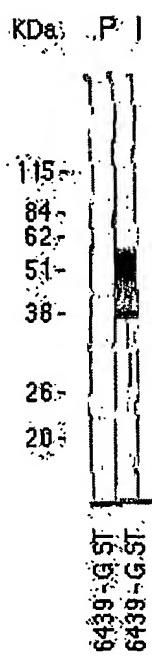
110/169

FIGURE 111**FIG. 111A****FIG. 111B**

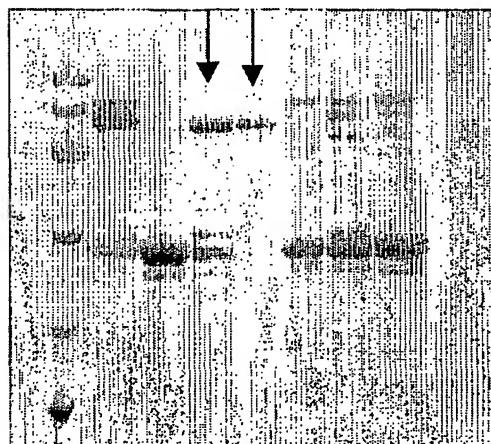
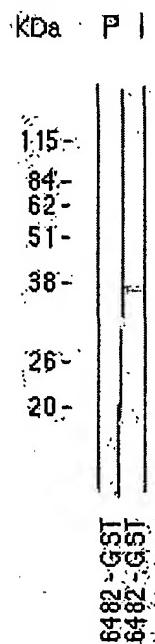
113/169

FIGURE 114**FIG. 114A****FIG. 114B**

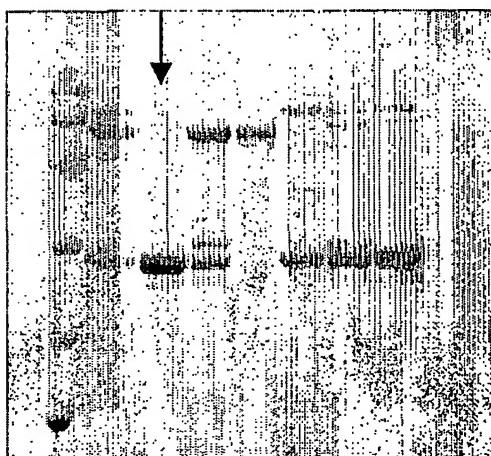
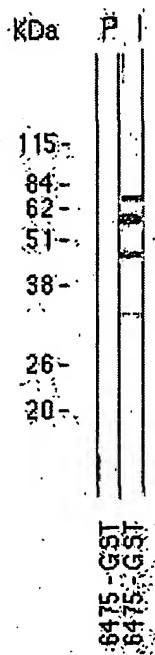
112/169

FIGURE 113**FIG. 113A****FIG. 113B**

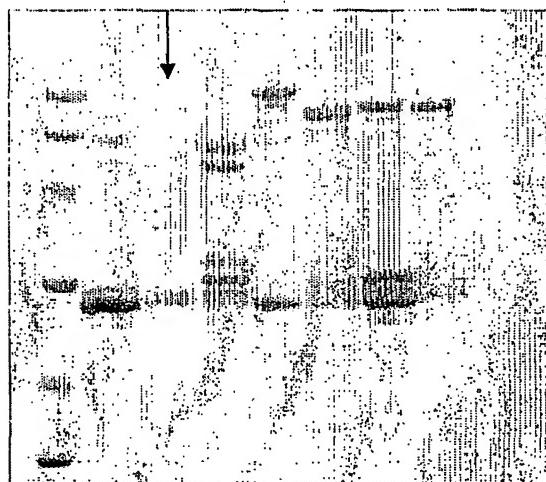
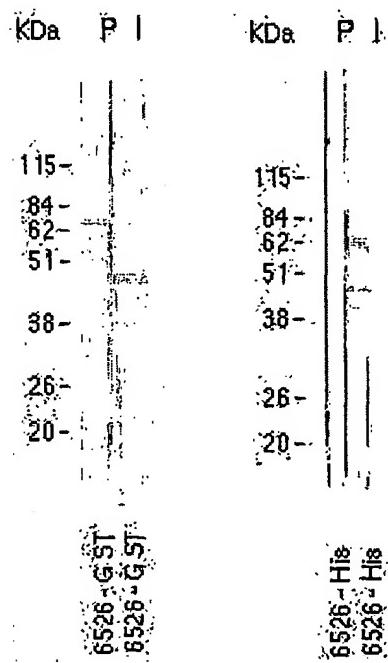
115/169

FIGURE 116**FIG. 116A****FIG. 116B**

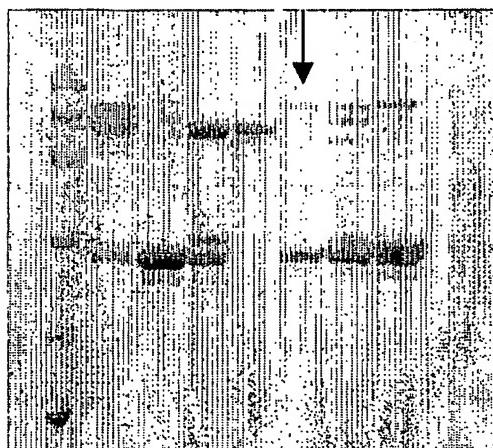
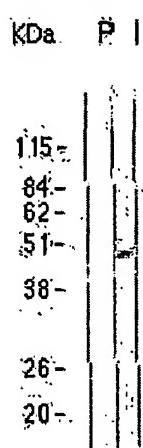
114/169

FIGURE 115**FIG. 115A****FIG. 115B**

117/169

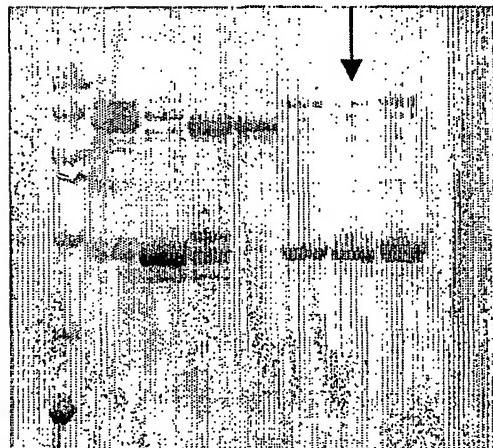
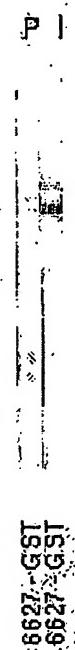
FIGURE 118**FIG. 118A****FIG. 118B**

116/169

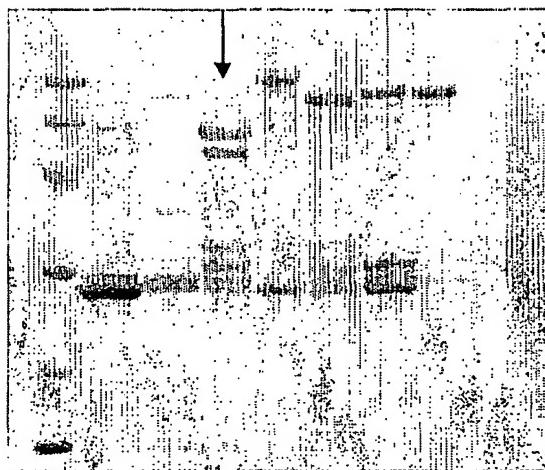
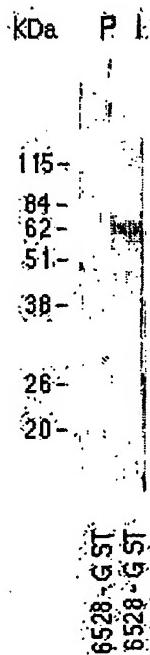
FIGURE 117**FIG. 117A****FIG. 117B**

6486-GST
6386-GST

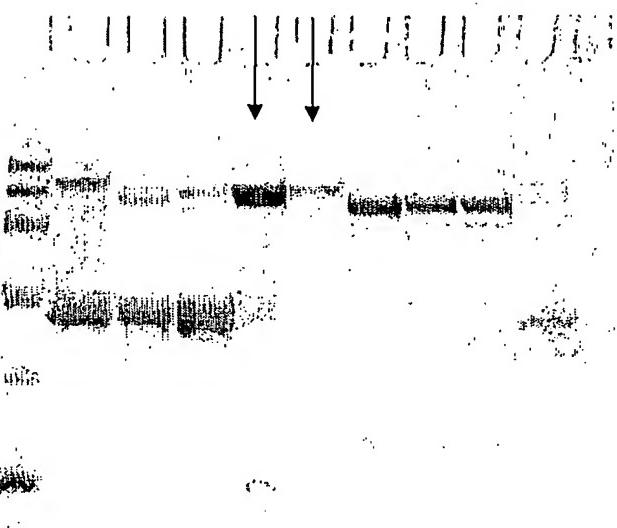
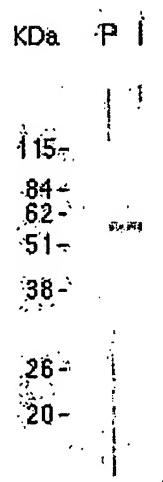
119/169

FIGURE 120**FIG. 120A****FIG. 120B**8627-GST
6927-GST

118/169

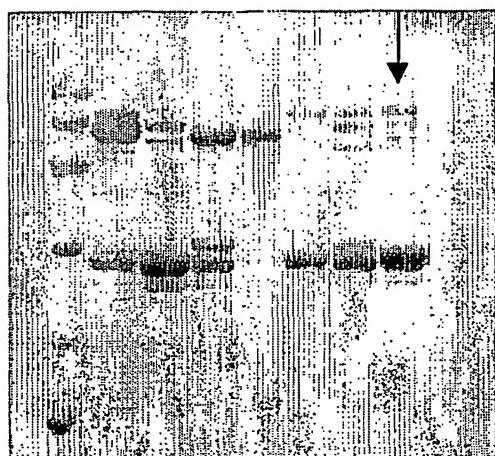
FIGURE 119**FIG. 119A****FIG. 119B**

121/169

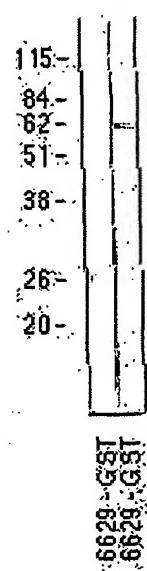
FIGURE 122**FIG. 122A****FIG. 122B**

6732-GST
6732-GST

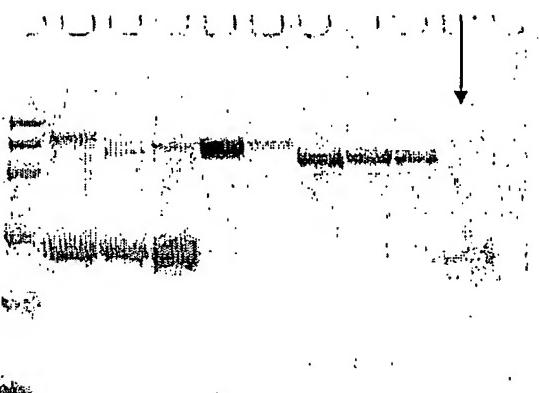
120/169

FIGURE 121**FIG. 121A**

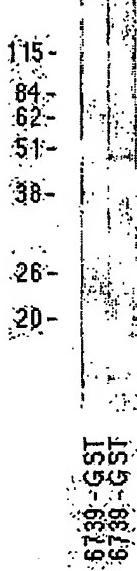
KDa P |

**FIG. 121B**

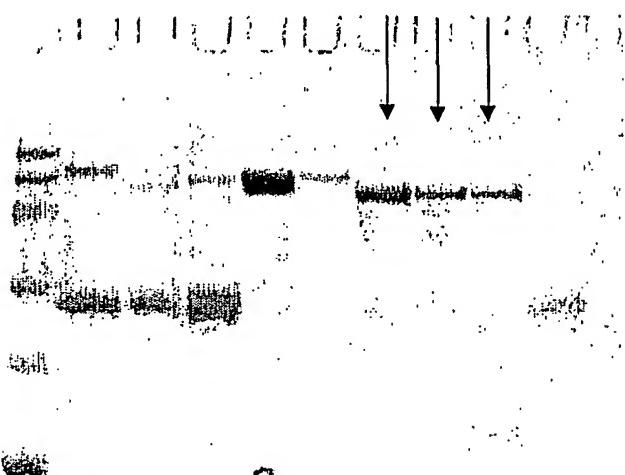
123/169

FIGURE 124**FIG. 124A**

KDa P |

**FIG. 124B**

122/169

FIGURE 123**FIG. 123A**

kDa P1

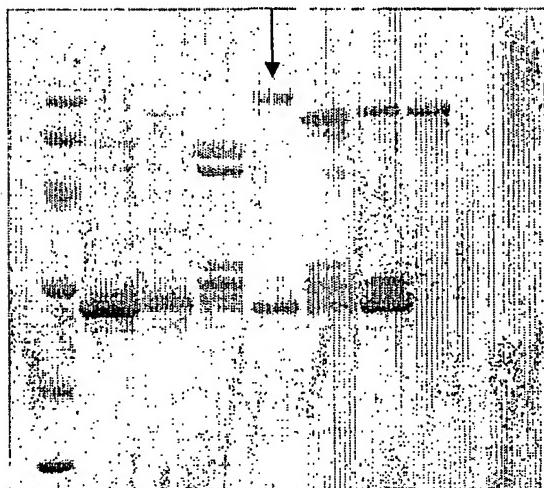
115-
84-
62-
51-
38-

26-
20-

FIG. 123B

6738-GST
6738-GST

125/169

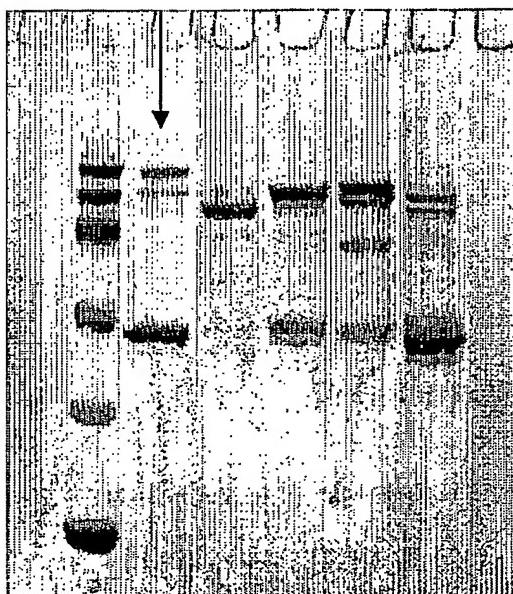
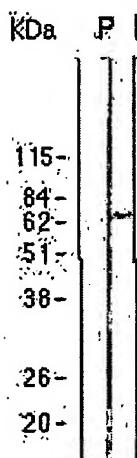
FIGURE 126**FIG. 126A**

kDa P F

115-
84-
62-
51-
38-26-
20-

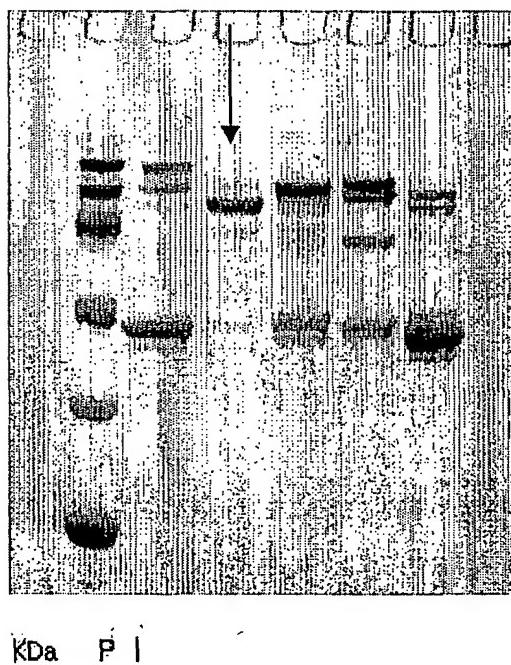
6742-GST
6742-GST**FIG. 126B**

124/169

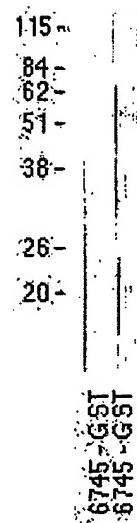
FIGURE 125**FIG. 125A****FIG. 125B**

GST
GST
6741-6741

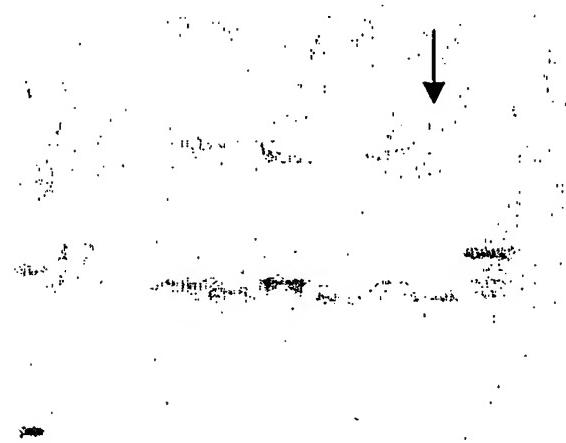
127/169

FIGURE 128**FIG. 128A**

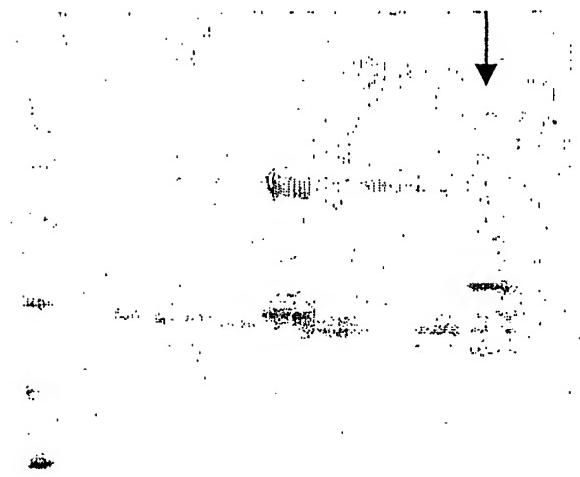
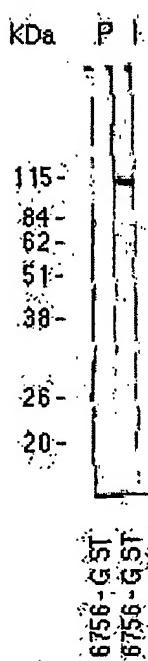
KDa P I

FIG. 128B6745-GST
6745-GST
6745-GST

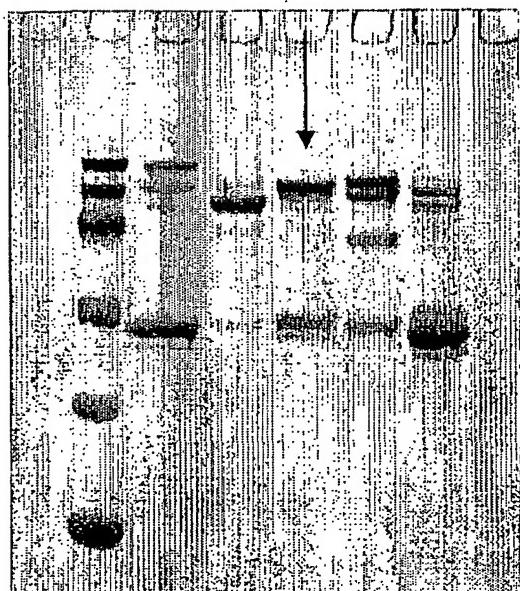
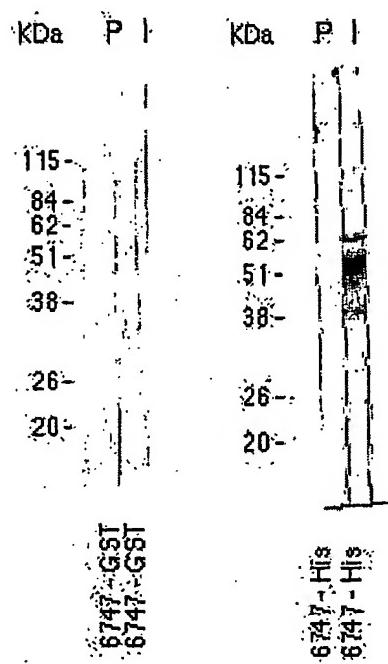
126/169

FIGURE 127**FIG. 127A****FIG. 127B**

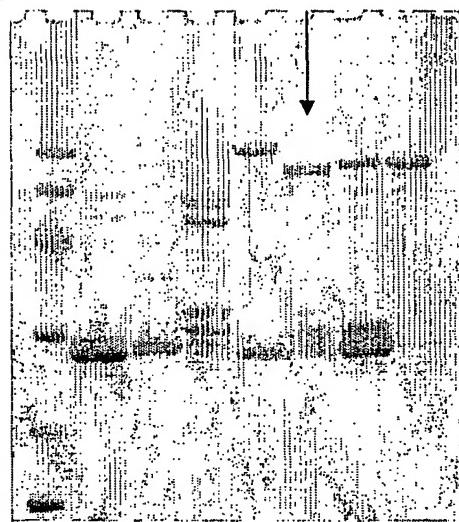
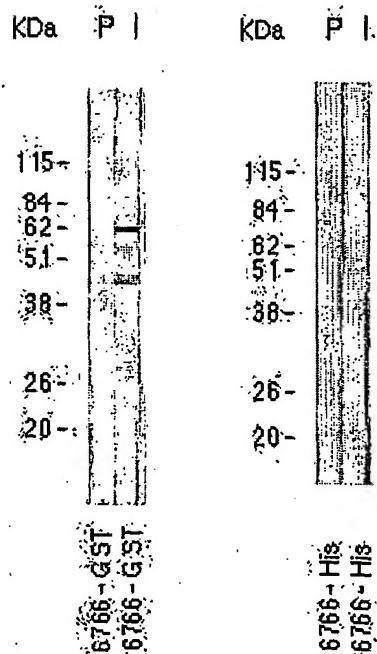
129/169

FIGURE 130**FIG. 130A****FIG. 130B**

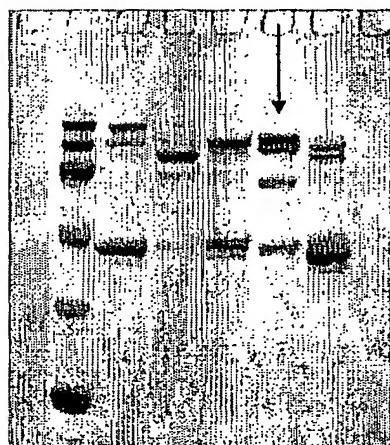
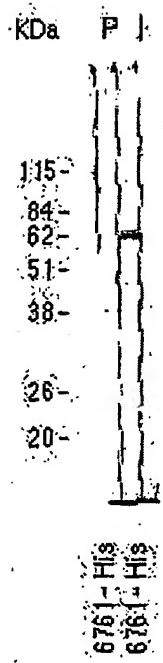
128/169

FIGURE 129**FIG. 129A****FIG. 129B**

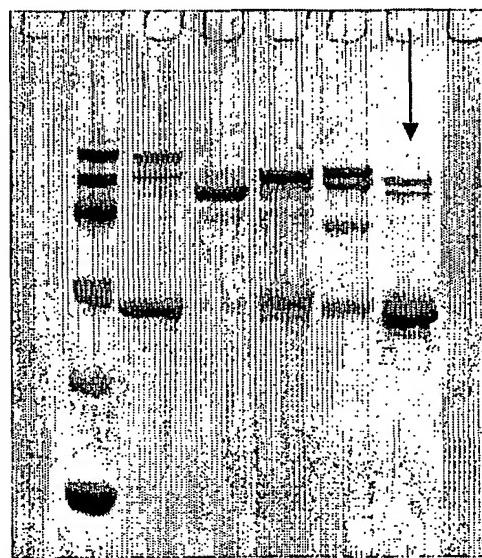
131/169

FIGURE 132**FIG. 132A****FIG. 132B**

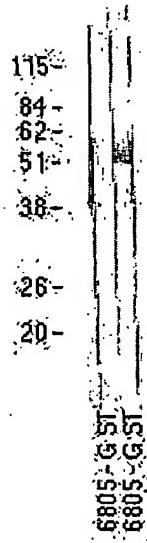
130/169

FIGURE 131**FIG. 131A****FIG. 131B**

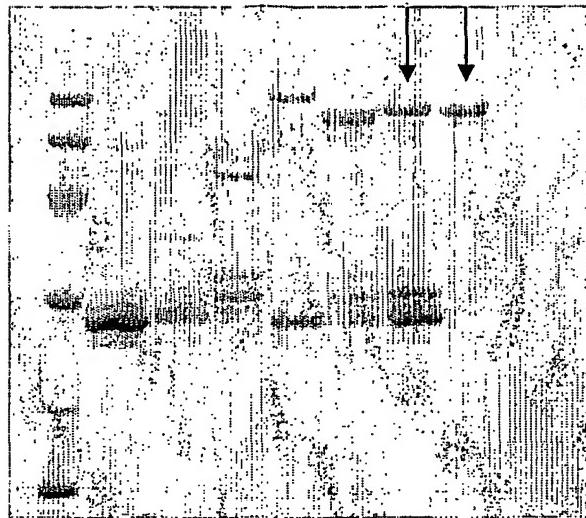
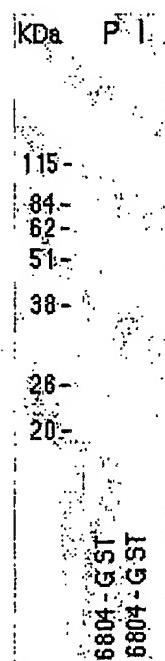
133/169

FIGURE 134**FIG. 134A**

kDa P:

**FIG. 134B**

132/169

FIGURE 133**FIG. 133A****FIG. 133B**

135/169

FIGURE 136

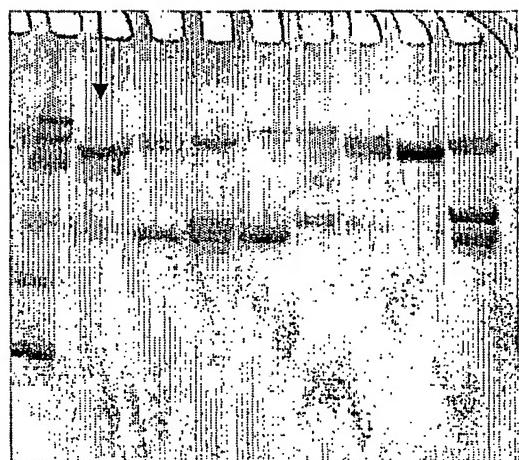


FIG. 136A

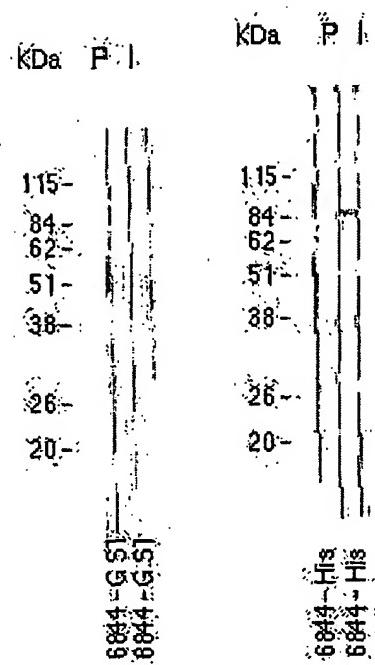
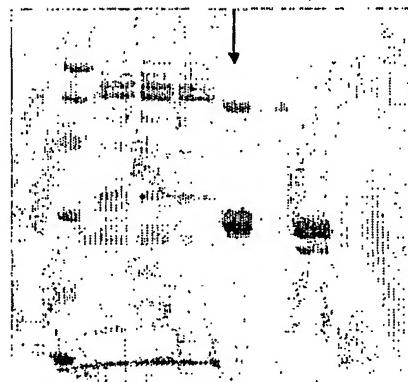
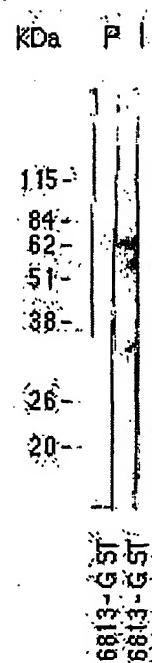
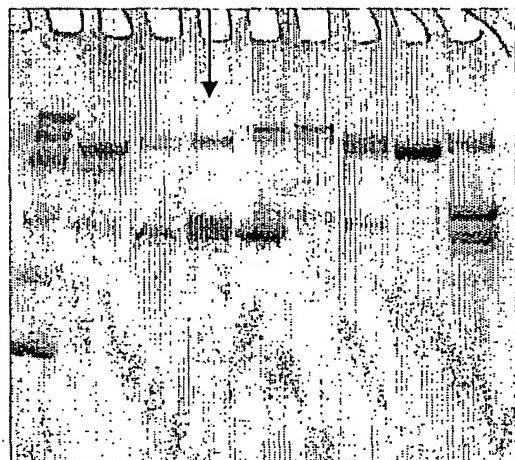
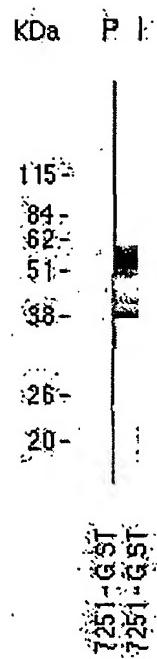


FIG. 136B

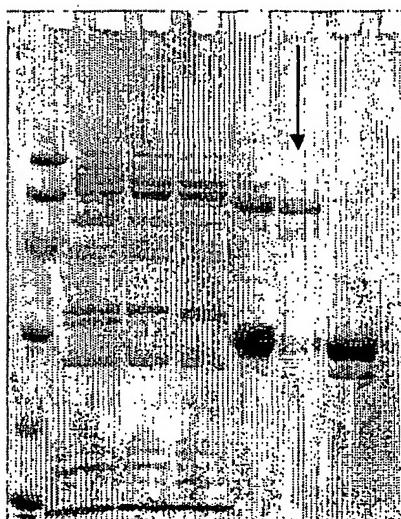
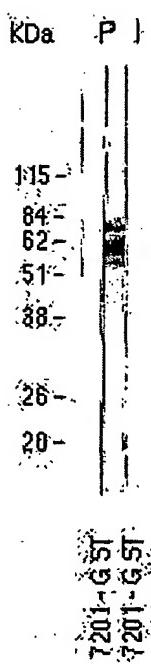
134/169

FIGURE 135**FIG. 135A****FIG. 135B**

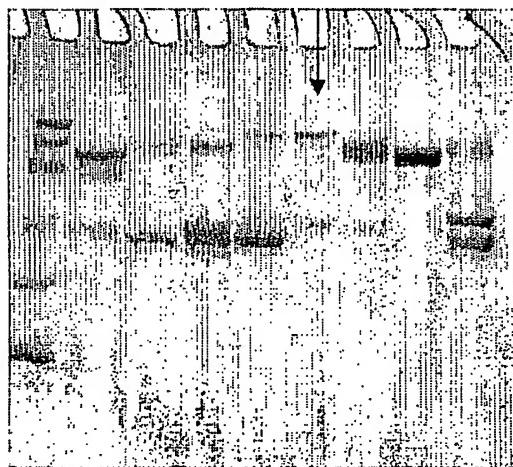
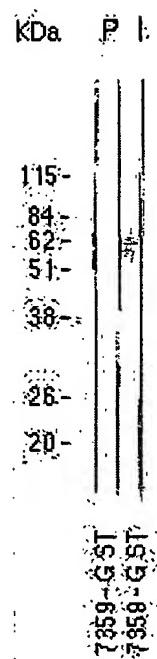
137/169

FIGURE 138**FIG. 138A****FIG. 138B**

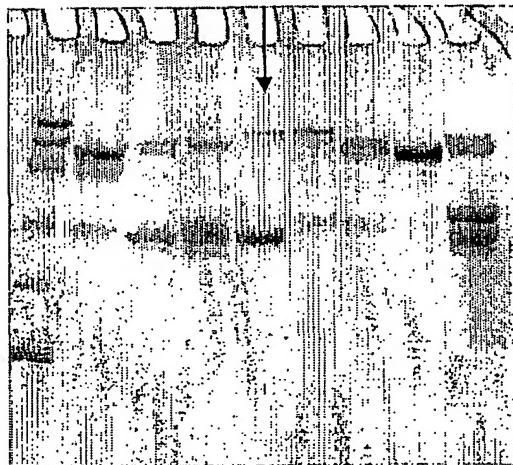
136/169

FIGURE 137**FIG. 137A****FIG. 137B**

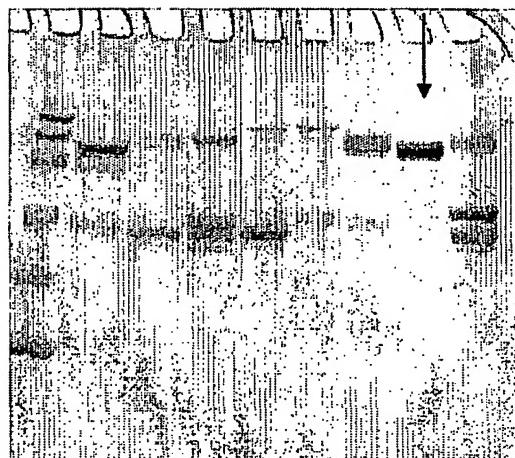
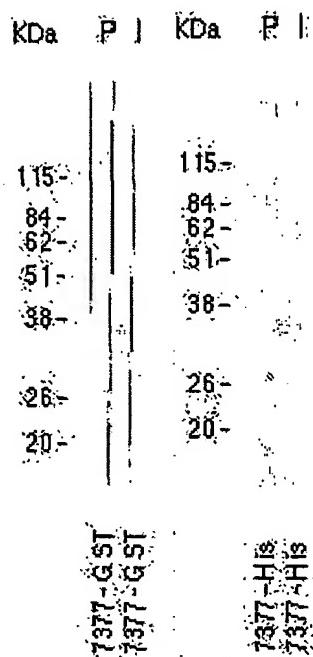
139/169

FIGURE 140**FIG. 140A****FIG. 140B**

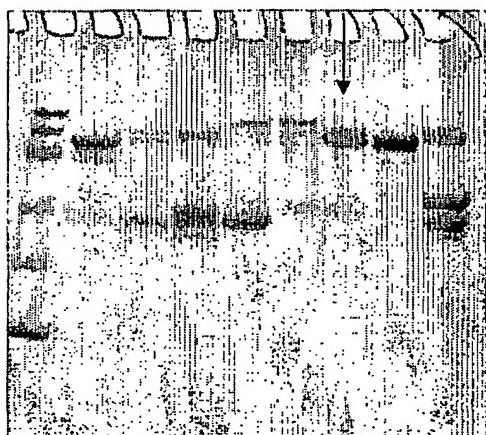
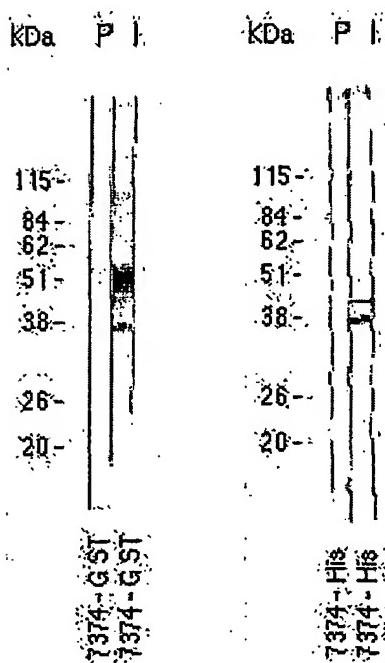
138/169

FIGURE 139**FIG. 139A****FIG. 139B**

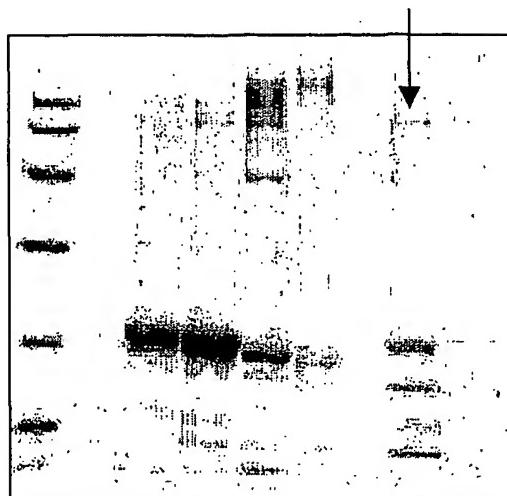
141/169

FIGURE 142**FIG. 142A****FIG. 142B**

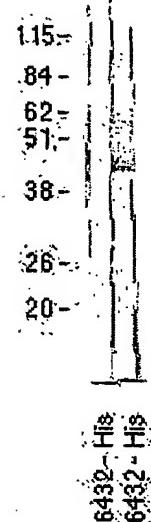
140/169

FIGURE 141**FIG. 141A****FIG. 141B**

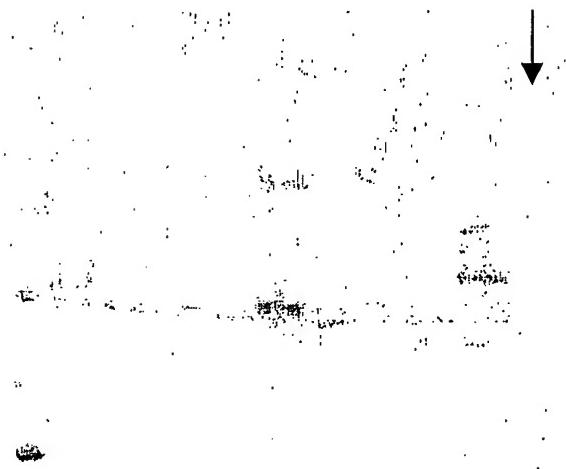
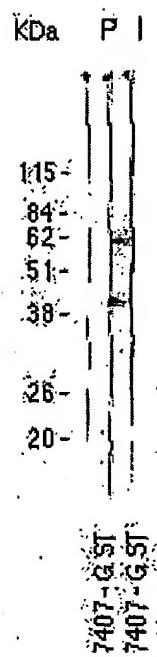
143/169

FIGURE 144**FIG. 144A**

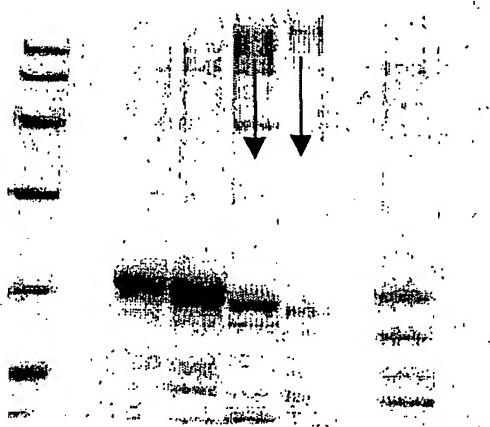
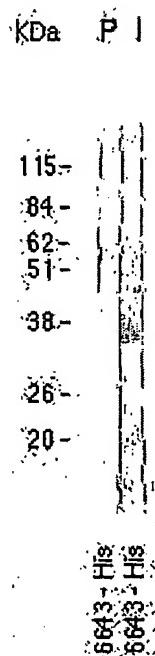
KDa P

**FIG. 144B**

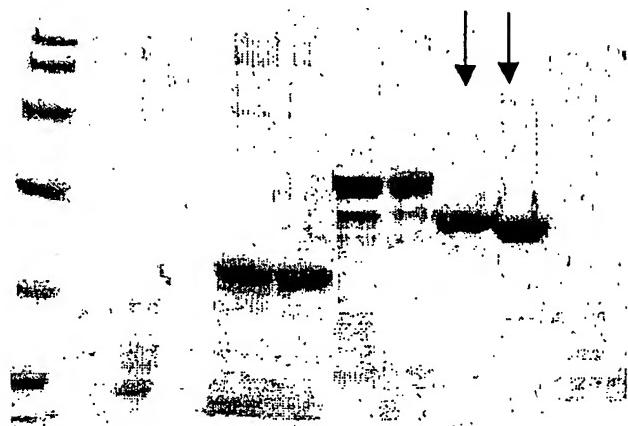
142/169

FIGURE 143**FIG. 143A****FIG. 143B**

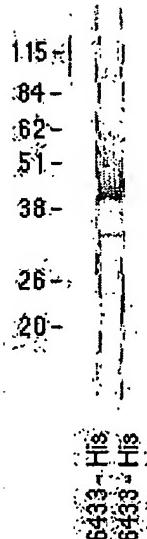
145/169

FIGURE 146**FIG. 146A****FIG. 146B**

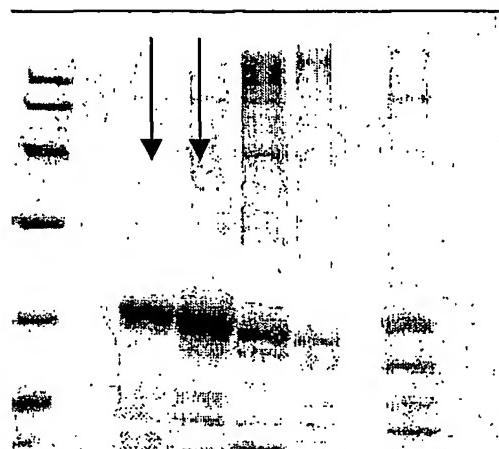
144/169

FIGURE 145**FIG. 145A**

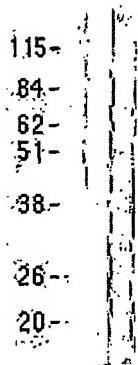
kDa P |

**FIG. 145B**6433 HS
6433 HS

147/169

FIGURE 148**FIG. 148A**

KDa P |

**FIG. 148B**Hs
7253 Hs
7253

146/169

FIGURE 147

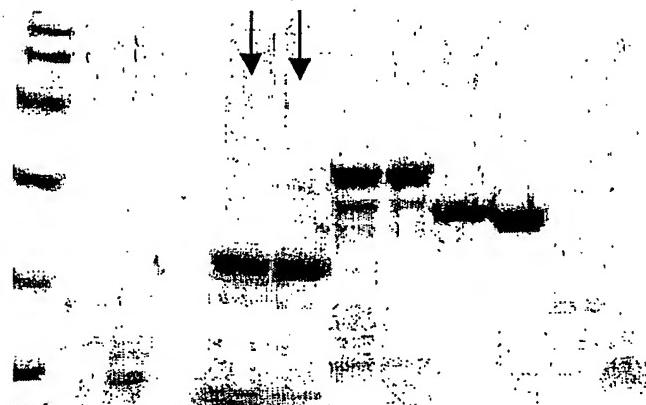


FIG. 147A

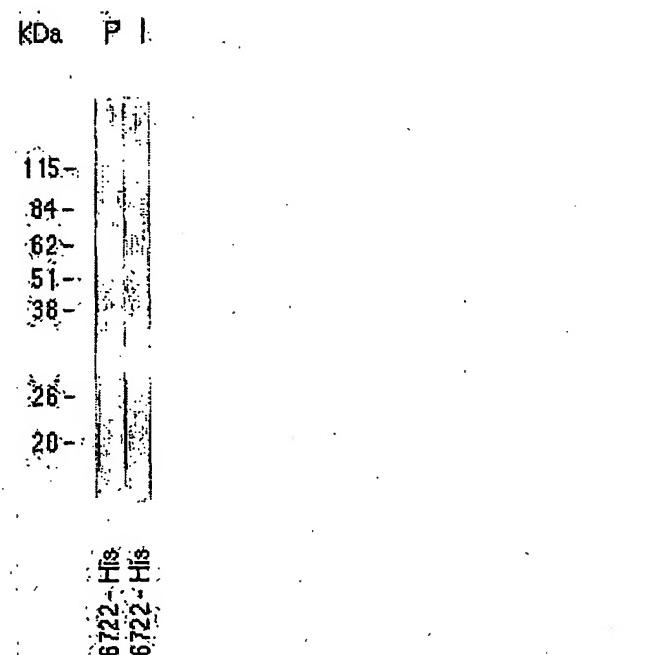
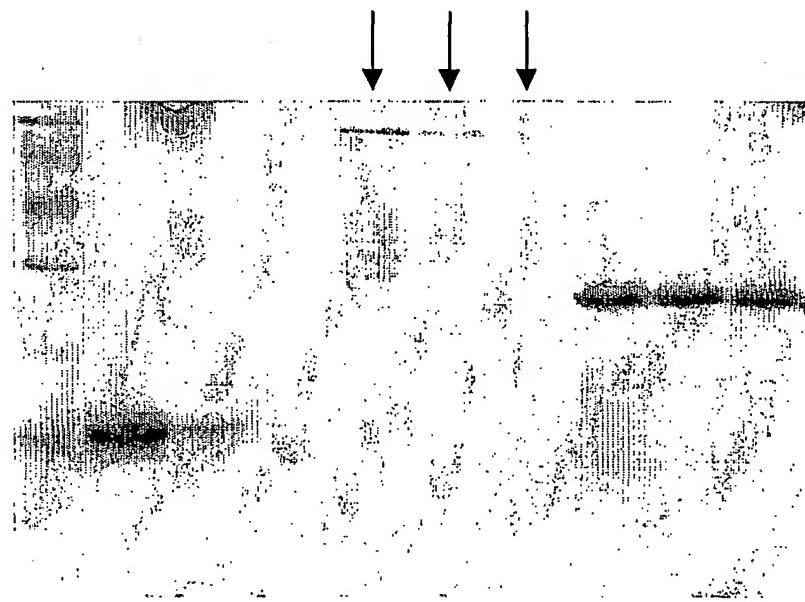
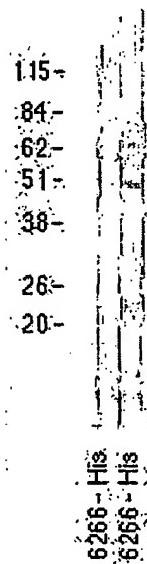


FIG. 147B

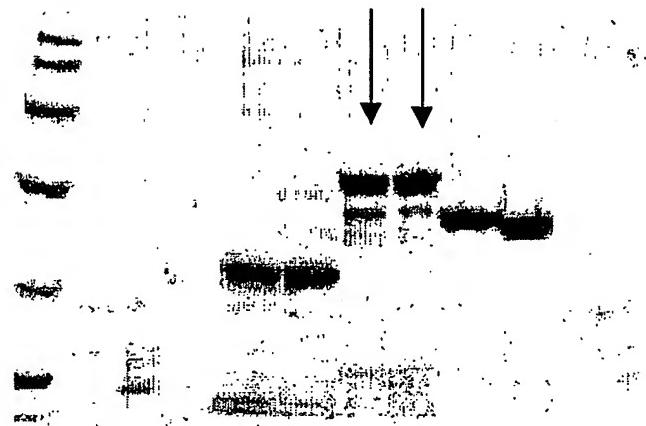
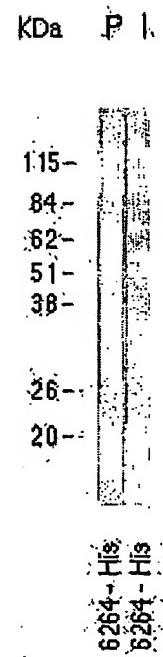
149/169

FIGURE 150**FIG. 150A****FIG. 150B**

KDa P.I.



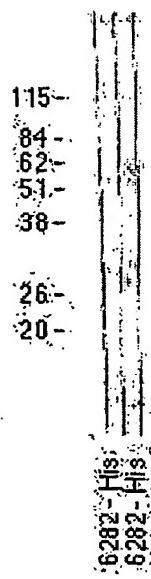
148/169

FIGURE 149**FIG. 149A****FIG. 149B**

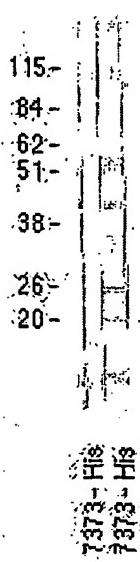
151/169

FIGURE 152**FIG. 152A****FIG. 152B**

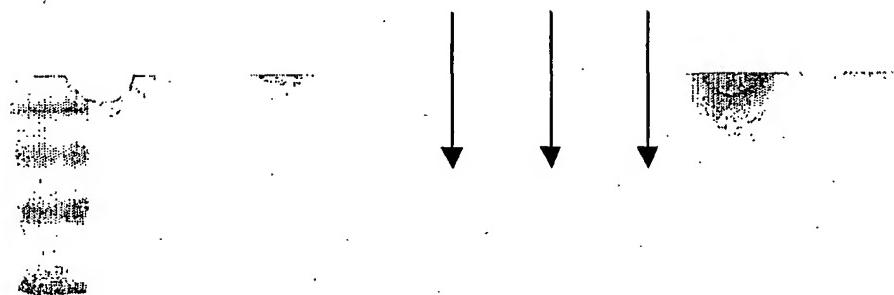
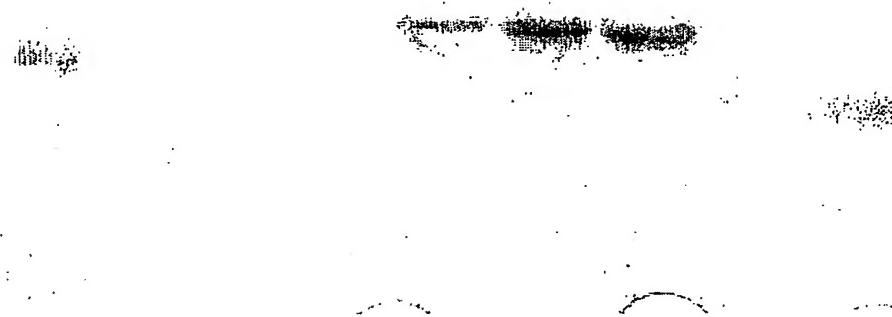
kDa P I

**FIGURE 153**

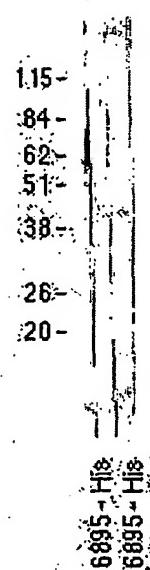
kDa P I



150/169

FIGURE 151**FIG. 151A****FIG. 151B**

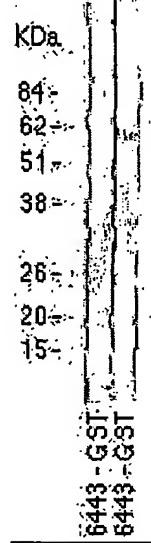
KDa P 1



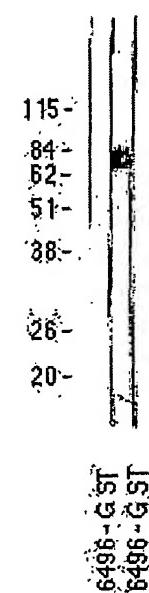
153/169

FIGURE 156

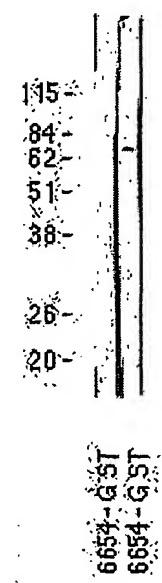
P I

**FIGURE 157**

KDa P I

**FIGURE 158**

KDa P I



152/169

FIGURE 154

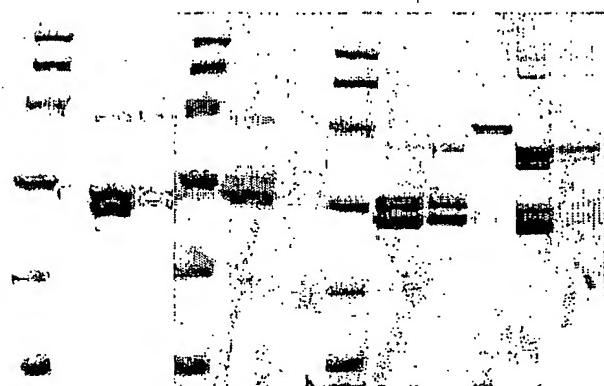


FIG. 154A

FIG. 154B

KDa P.I.

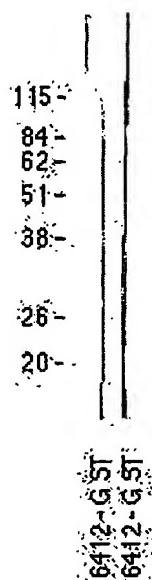
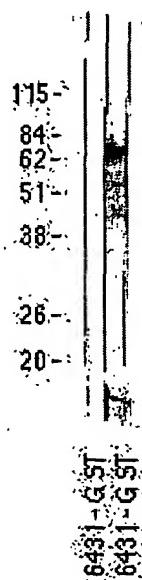


FIGURE 155

kDa P I



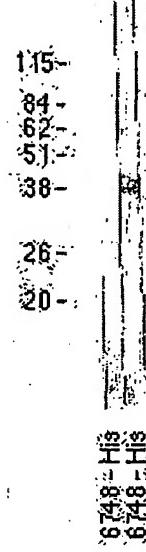
155/169

FIGURE 161**FIG. 161A****FIG. 161B**

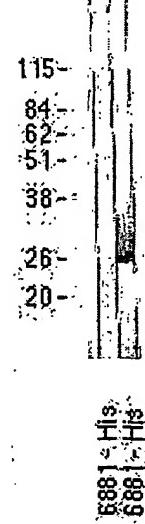
kDa P1

**FIGURE 162**

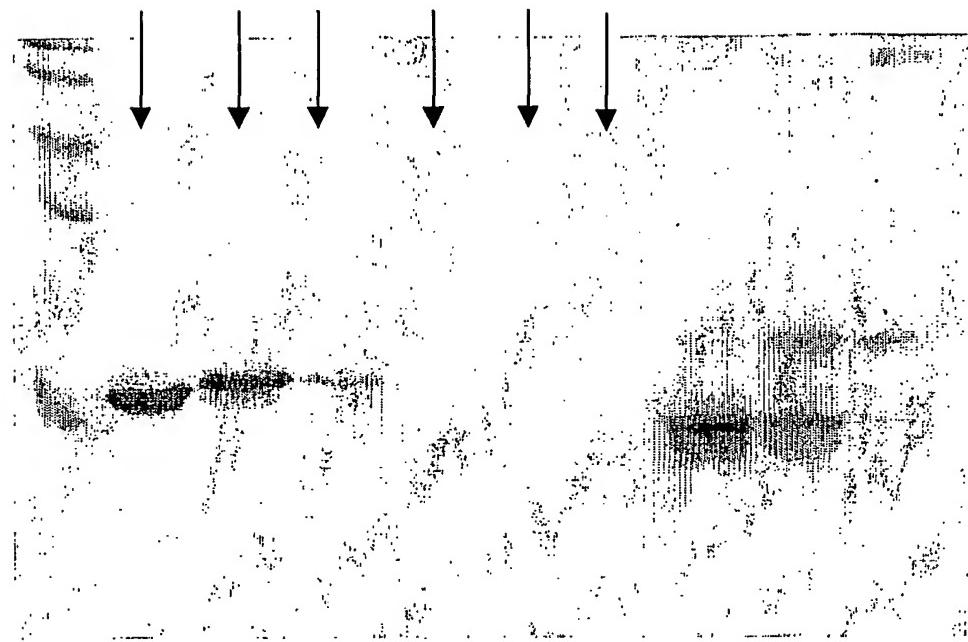
kDa P1

**FIGURE 163**

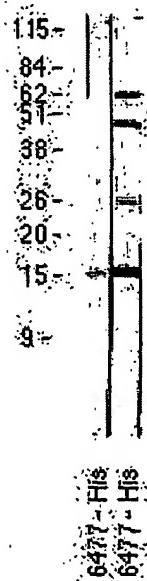
kDa P1



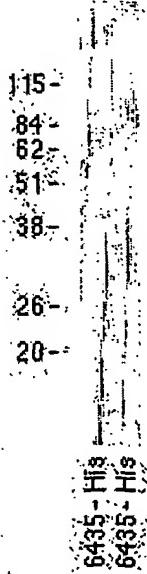
154/169

FIGURE 159**FIG. 159A****FIG. 159B**

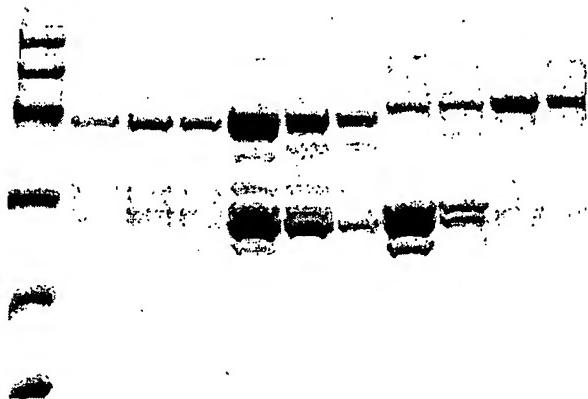
KDa P I

**FIGURE 160**

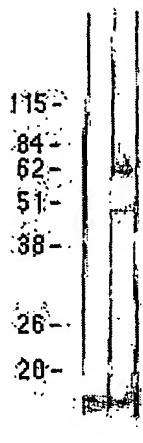
KDa P I



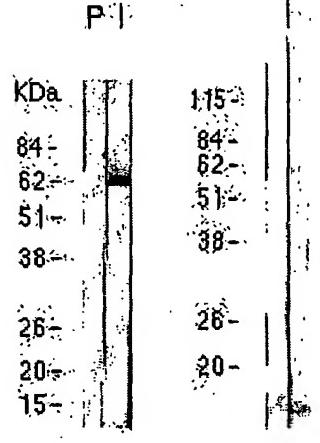
157/169

FIGURE 167**FIG. 167A****FIG. 167B**

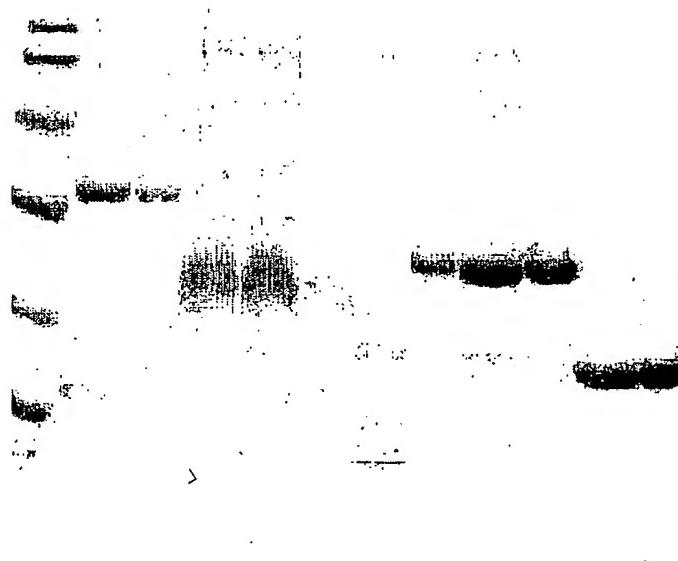
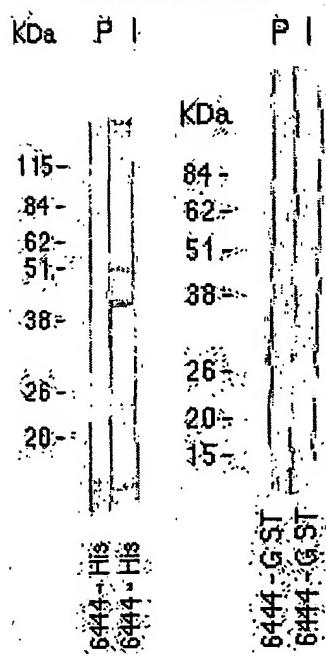
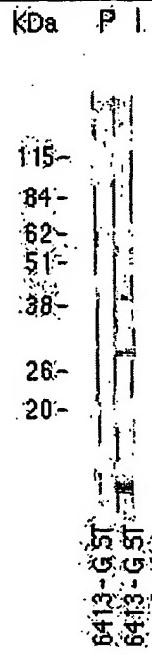
KDa P I

6540-His
6540-GST**FIGURE 168**

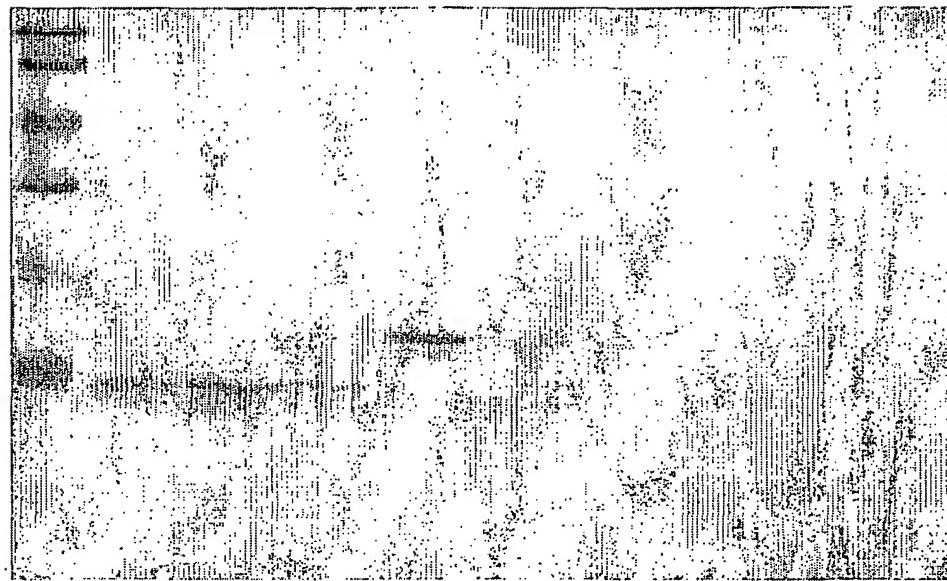
KDa P I

6540-His
6540-GST

156/169

FIGURE 164**FIG. 164A****FIG. 164B****FIGURE 165****FIGURE 166**

159/169

FIGURE 171**FIG. 171A****FIG. 171B**

KDa R I

115-
84-
62-
51-
38-
26-
20-

662 His
662 His

FIGURE 172

KDa R I

115-
84-
62-
51-
38-
26-
20-

674 His
674 His

FIGURE 173

KDa P I

115-
84-
62-
51-
38-
26-
20-
15-
9-

6497 His
6497 His

158/169

FIGURE 169

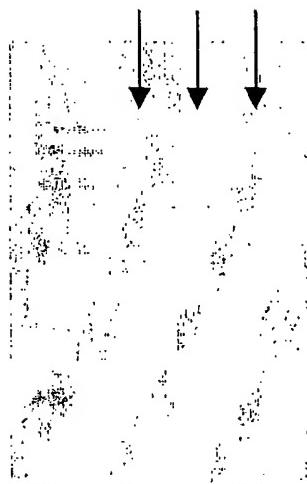
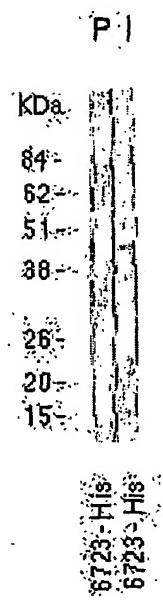
KDa P I

115
84
62
51
38
26
206743-G5
6743-G5**FIGURE 170**

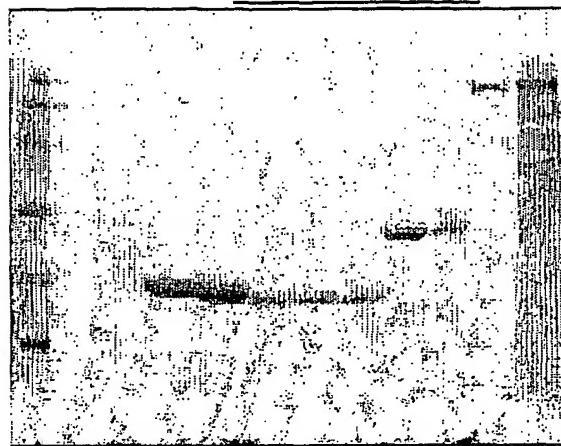
KDa P I

115
84
62
51
38
26
207041-G5
7041-G5

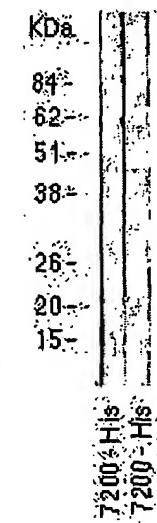
161/169

FIGURE 179**FIG. 179A****FIG. 179B**

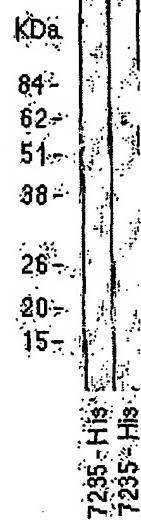
160/169

FIGURE 174**FIG. 174A****FIG. 174B**

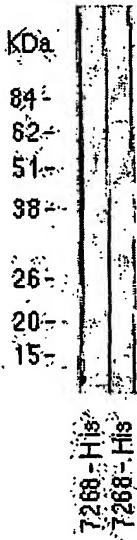
P.I.

**FIGURE 175**

P.I.

**FIGURE 176**

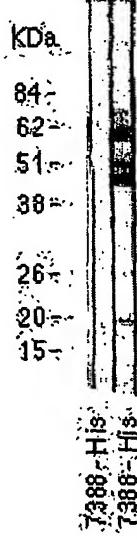
P.I.

**FIGURE 177**

P.I.

**FIGURE 178**

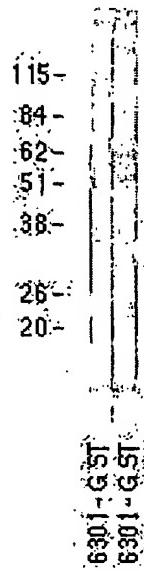
P.I.



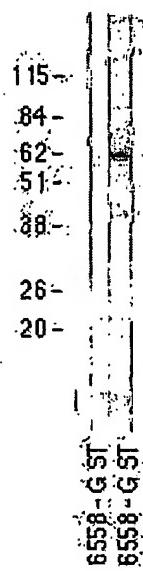
163/169

FIGURE 181

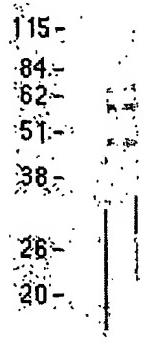
KDa P I

**FIGURE 182**

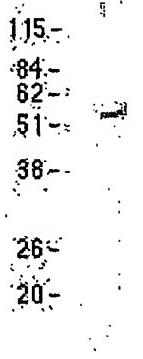
KDa P I

**FIGURE 183**

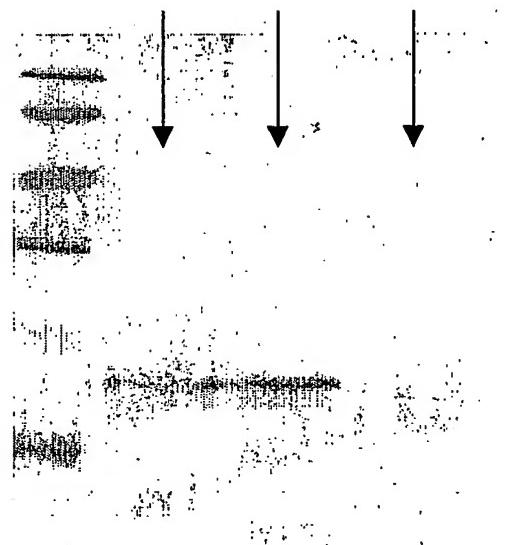
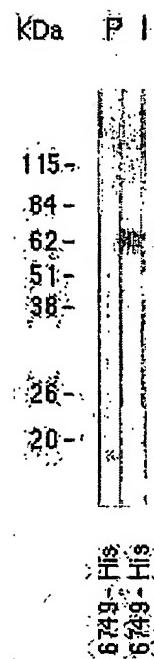
KDa P I

**FIGURE 184**

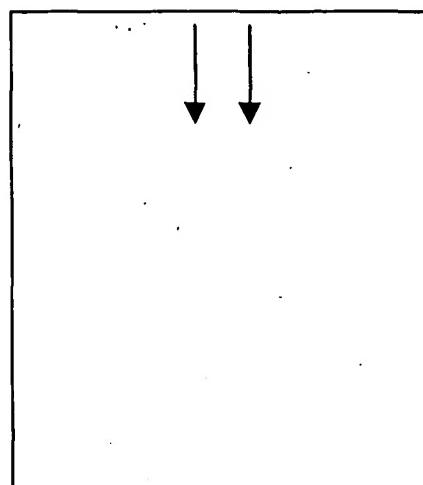
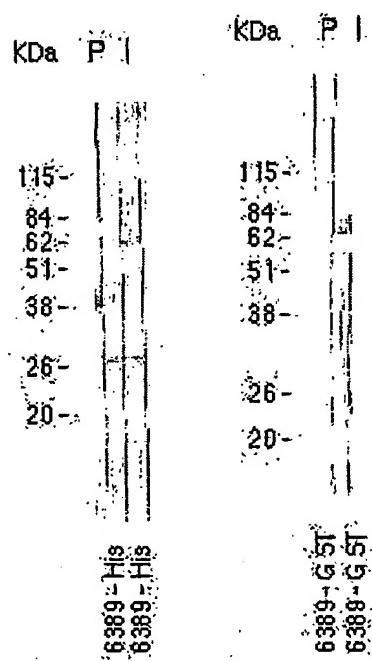
KDa P I



162/169

FIGURE 180**FIG. 180A****FIG. 180B**

165/169

FIGURE 186**FIG. 186A****FIG. 186B**

164/169

FIGURE 185

KDa P I

115-

84-

62-

51-

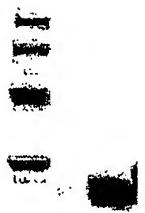
38-

26-

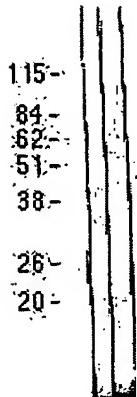
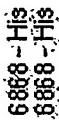
20-

GST
GST
6642 6642
6643 6643

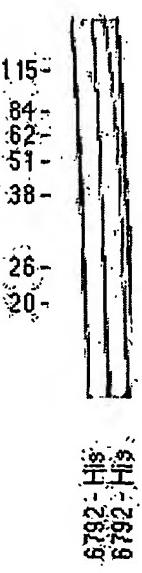
167/169

FIGURE 188**FIG. 188A**

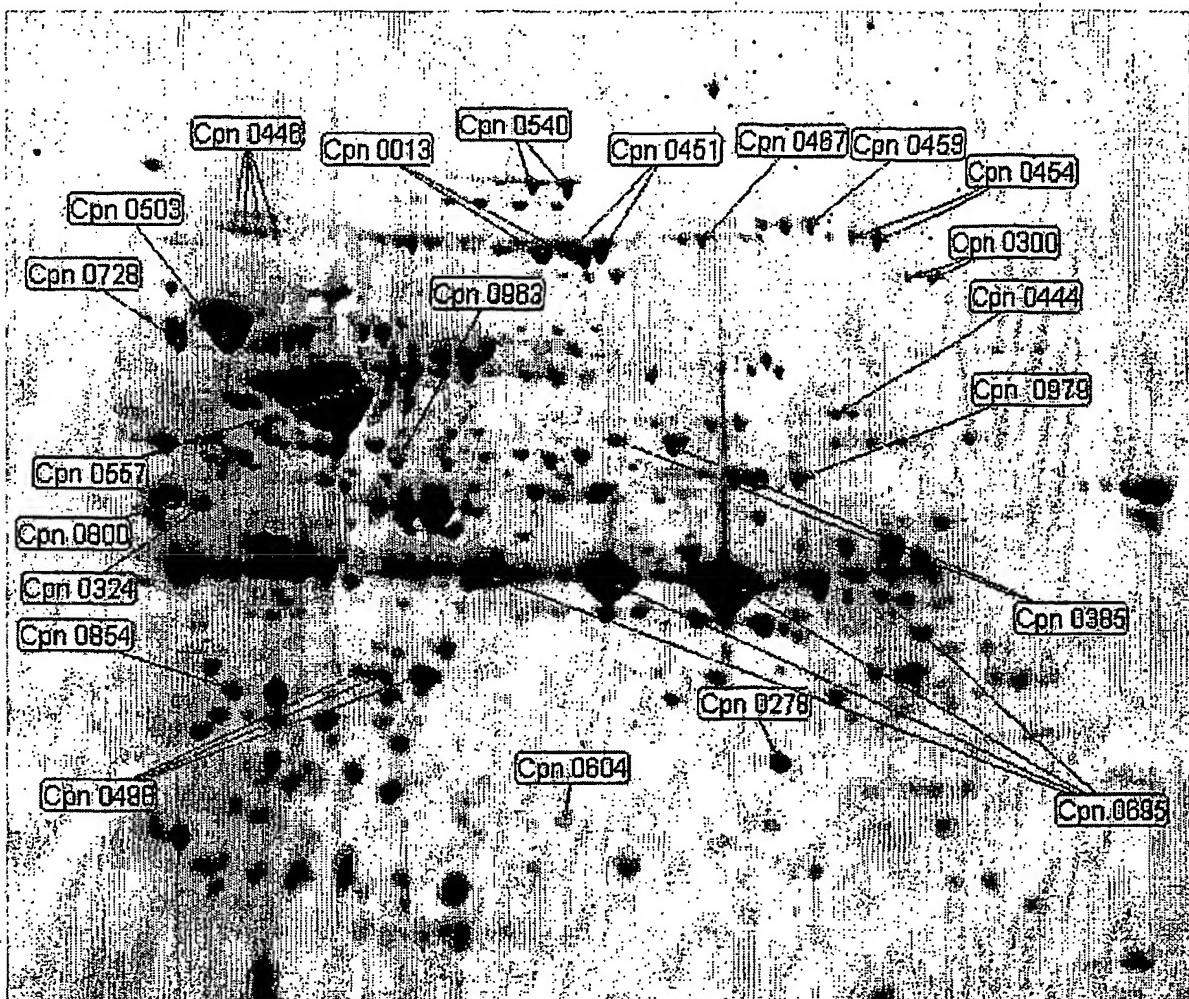
kDa P.I.

**FIG. 188B**

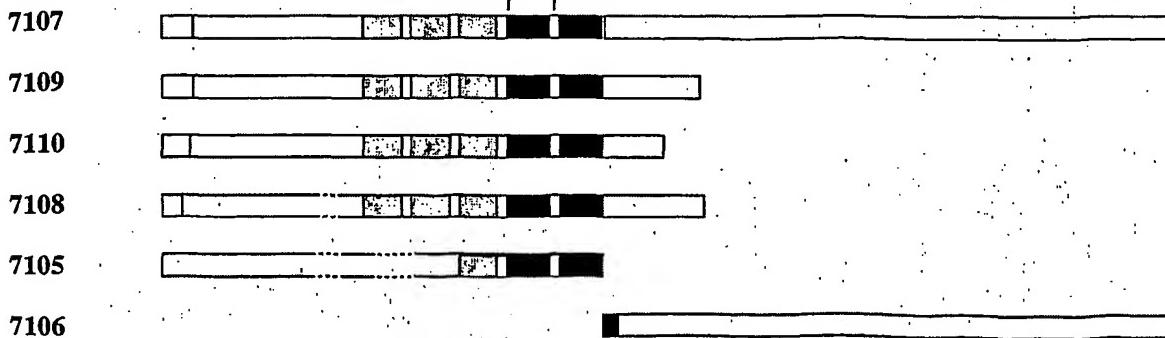
166/169

FIGURE 187**FIG. 187A****FIG. 187B**

169/169

FIGURE 190**FIGURE 191**

SVIVG . VSTNSEHRYHAFQYADGQMVDLCTMIGCPESYAQGVSGDCK
 KVIVVG . HSTRTDGEYRAFKYVDGRMIDLCTMIGGSASFAFGVSSDGK
 KVIVVG . RSETYYGEVHAFCHKNGVMSDLGTLGGSYSAAKGVSATGK
 KVIVVG . WSTTNNGETHAFMHKDETMHDLCMIGCGGFSVATGVSAADR
 TIVVGSMESTITRKTTAVKWVNNVPTYLCTLCGDASTGLYISGDGT



168/169

FIGURE 189**FIG. 189A**

kDa P L

115-
84-
62-
51-
38-
26-
20-115
84
62
51
38
26
20**FIG. 189B**